Available online at www.sciencedirect.com
SCIENCE DIRECT
Talanta 68 (2005) 382–387
Talanta
www.elsevier.com/locate/talanta

Usefulness of chemometrics and mass spectrometry-based electronic nose to classify Australian white wines by their varietal origin

Daniel Cozzolino [a,b,*], Heather E. Smyth [a,b,c,1], Wies Cynkar [a,b,1], Robert G. Dambergs [a,b,1], Mark Gishen [a,b,1]

[a] *The Australian Wine Research Institute, Chemistry Department, Waite Road, Urrbrae, P.O. Box 197, Glen Osmond - Urrbrae Adelaide, Adelaide, SA 5064, Australia*
[b] *The Cooperative Research Centre for Viticulture, P.O. Box 154, Glen Osmond, SA 5064, Australia*
[c] *School of Agriculture and Wine, Faculty of Sciences, The University of Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia*

Available online 3 October 2005

Abstract

A combination of mass spectrometry-based electronic nose (MS e_nose) and chemometrics was explored to classify two Australian white wines according to their varietal origin namely Riesling and unwooded Chardonnay. The MS e_nose data were analysed using principal components analysis (PCA), discriminant partial least squares (DPLS) and linear discriminant analysis (LDA) applied to principal components scores and validated using full cross validation (leave one out). DPLS gave the highest levels of correct classification for both varieties (>90%). LDA classified correctly 73% of unwooded Chardonnay and 82% of Riesling wines. Even though the conventional analysis provides fundamental information about the volatile compounds present in the wine, the MS e_nose method has a series of advantages over conventional analytical techniques due to simplicity of the sample-preparation and reduced time of analysis and might be considered as a more convenient choice for routine process control in an industrial environment. The work reported here is a feasibility study and requires further development with considerably more commercial samples of different varieties. Further studies are needed in order to improve the calibration specificity, accuracy and robustness, and to extend the discrimination to other wine varieties or blends.

Keywords: MS e_nose; Mass spectrometry-based electronic nose; Chemometrics; Classification; White wines

1. Introduction

The main challenge of any quality-control system to be used in a modern and competitive food industry is to replace time-consuming laboratory analysis used in both quality monitoring and process control, with relatively fast and cheaper measurements suitable for routine operation. The application of chemometrics techniques to mass spectrometry (MS)-based electronic nose (e_nose) data has been investigated by several authors as a means of differentiating food samples on the basis of both aroma and volatile compounds in the food industry [1]. Food product characterisation based on the analysis of their aroma properties is a widely used technique [2–3]. Nowadays, analytical solutions for food composition imply the use of gas chromatography–mass spectrometry (GC–MS) techniques, but analysis can be time-consuming, due to sample-preparation steps and complex data interpretation. Recent research has shown that rapid analysis of volatile fractions by MS without chromatographic separation produces signals containing useful information that can be used to produce a fingerprint of any given food based on its aroma profile [4–5]. Few studies have examined the use of electronic noses or gas sensors to characterise the aroma of wine [1,6], mainly because major compounds in the samples headspace, such as ethanol, cause interference with the gas sensor [1]. This limitation does not exist with MS e_nose where the headspace is monitored and the whole spectra are analysed [1,6]. Therefore, MS e_nose has a great potential to be used for monitoring the quality of wine and other alcoholic beverages [1,5–8]. The MS e_nose and gas sensor array techniques have been applied to the classification of foods (fish and meat), beverages (wine, beer

Abbreviations: DPLS, discriminant partial least squares; GC–MS, gas chromatography–mass spectrometry; LDA, linear discriminant analysis; MS e_nose, mass spectrometry electronic nose; PCA, principal component analysis; PCs, principal components; Vis-NIR, visible and near infrared

* Corresponding author. Fax: +61 8 8303 6601/9449.
E-mail address: Daniel.Cozzolino@awri.com.au (D. Cozzolino).
[1] Fax: +61 8 8303 6601/9449.

0039-9140/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2005.08.057

and spirits) and have also been used for the quality control of industrial products such as paper [1,5–8]. Most of the references in the literature on the application of such techniques in wine are related to monitoring aroma and other volatile compounds in either ferments or in the final wine product [9,12–14]. For example, the use of MS e_nose was explored as a rapid technique for fingerprinting of volatiles in wines before and after malolactic fermentation, in wine stored in oak barrel types and in Semillon wine bottled with different closures [14].

Multivariate methods and chemometrics have been used to interpret and extract information from complex data obtained by instrumental techniques in the analysis of foods and wine [2,10,15]. The application of multivariate statistical techniques such as principal component analysis (PCA) or discriminant analysis (e.g., linear discriminant analysis, LDA, or discriminant partial least squares, DPLS) provides the possibility to use and understand the data generated by instrumental techniques based on the overall properties of the sample and perform a classification without the need for additional compositional chemical data [2,15].

A single MS e_nose mass spectrum might provide means of characterising complex features of wine including aroma, stability (e.g., protein or heat stability), oxidation, quality grading or blending. In addition, it could assist in determining the relationship between the chemical composition and sensory characteristics of wine. Despite the large amount of research carried out to date to attempt to differentiate between varieties of wine, there is only limited information published concerning the use of chemometrics on headspace gas sensor data to differentiate wine samples on the basis of their variety [11,12,14].

The study presented here is a part of the ongoing evaluation of rapid instrumental methods being carried out by The Australian Wine Research Institute in order to adapt competitive and modern instrumental techniques for the Australian wine industry.

The aim of this work was to investigate the potential of MS e_nose as a rapid and low-cost technique to discriminate between two commercial white wine varieties available in Australia (i.e., Riesling and unwooded Chardonnay) using the mass spectra of their volatile constituents without time-consuming analysis of chemical composition by GC–MS.

2. Materials and methods

2.1. Wine samples

A total of 150 white wine samples of two varieties, namely Riesling and unwooded Chardonnay, were analysed. Of the total, 120 samples were selected from an experiment that was part of a broader wine flavour study [16,31] with three replicate bottles of each (same vintage, label and closure) of 20 individual commercial Riesling or unwooded Chardonnay wine labels. Additional 30 samples of Riesling were sourced from a consumer sensory experiment. The wines selected were chosen from a larger sample set by a series of preliminary informal sensory assessments, with the primary criteria for selection being that the wines should exhibit the broadest possible range of sensory properties within each variety. All samples were commercially available at the time and were generally representative of the range of retail prices, regions, and vintages of Riesling and unwooded Chardonnay wines available in the Australian market. Of the Riesling wines, vintages comprised 2002, 2001, 1999, 1997, 1996 and 1993. The Chardonnay wines were all from 2002 except for three samples made in the 2001 vintage. Summary information of the chemical composition of the wines of each variety is detailed in previous reports [16–17].

2.2. Mass spectrometry electronic nose (MS e_nose)

To measure the volatile patterns of the two wine varieties, 5 mL of wine were sealed in 10 mL headspace vials. Samples were analysed on the *Chemical Sensor* (HP 4440, Hewlett Packard) equipped with headspace sampler (HP 7694, Model G 1290 A) [18–20]. The experimental conditions of the headspace sampler were as follows: oven temperature 75 °C, loop temperature 90 °C, transfer line temperature 95 °C, vial equilibration time 20 min, headspace cycle time 4.2 min, pressurise time 0.3 min, loop fill time 0.15 min, loop equilibration time 0.02 min, and injection time 0.5 min. The carrier gas was helium, pressure 4.2 psi and the vial was pressurized at 14 psi. The total analysis time per sample was approximately 25 min. The components in the headspace of the vials were passed directly to the mass detector without any chromatographic separation or sample pretreatment. In this way, for any given measurement, the resulting mass spectrum gives a fingerprint of the wine volatiles. Positive ion electron impact spectra at 70 eV were recorded in the range m/z 50–180. Additionally, in each MS e_nose run, a solution of 12% ethanol was used as a marker to identify the ions related with ethanol. Data and instrument control was achieved with the *Pirouette* software [18–20]. Operational conditions and protocols used in this study were similar to those reported elsewhere [14].

2.3. Chemometrics

Data from MS e_nose were exported from the *Pirouette* software for chemometric analysis into *The Unscrambler* software (version 9.1, CAMO ASA, Norway). Principal component analysis (PCA), discriminant partial least squares (DPLS) and linear discriminant analysis (LDA) were performed with full cross validation [21–23]. Full cross validation (leave one out) was used to validate the models developed. The maximum number of factors (terms) in the PLS models were selected by the criterion of the lowest number of factors that gave the closest to minimum value of the PRESS (prediction residual error sum of squares) function in order to avoid overfitting of the models.

As a pre-treatment before both PCA and DPLS analysis, all data were centered [21–24]. Before performing principal component analysis, MS e_nose data were pre-processed in order to account for baseline effects, retention time drifts and variations in peak shapes between the samples analysed [21–24]. In this study, auto-scaling was performed by smoothing (moving average of each of seven data points) and mean normalisation provided by *The Unscrambler* software was used. The moving average reduced the noise and made it easier to observe the start

and end of the peaks in the spectra of each wine sample analysed. Mean normalisation is the most classical case of normalisation. It consists of dividing each row of a data matrix by its average, thus neutralising the influence of the hidden factors, such as noise or drift between analyses. It is equivalent to replacing the original variables by a profile value centred around 1. Only the relative values of the variables are used to describe the sample, and the information carried by their absolute level is dropped. This is appropriate as all variables are measured in the same unit, and their values are assumed to be proportional to a factor, which cannot be directly taken into account in the analysis [22]. For instance, this transformation is used in chromatography to express the results in the same units for all samples. This normalisation was used in this study because the samples were analysed on different days and to eliminate the effect of the differences in the alcohol concentrations of the samples.

Principal component analysis (PCA) is a well-known technique used for reducing the dimensionality of the data, detecting the number of components and visualising the outliers. It is one of the most commonly applied techniques in multivariate data analysis [21–24]. PCA is a mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired degree of accuracy [21–24]. In this study, PCA was used to derive the first 20 principal components from the MS e_nose data and to examine the possible grouping of samples.

The DPLS models were developed using a no metric dummy variable (set to 1 for Riesling, and 2 for unwooded Chardonnay) to test the ability of the method to discriminate between the two white wine varieties; this approach is often referred to as a discriminant PLS (DPLS) [16,21–24]. The small number of samples used prevented the development of a definitive prediction model for DPLS but was still sufficient to enable a preliminary assessment of the potential of this technique to classify the two wine varieties.

Linear discriminant analysis (LDA) like DPLS regression is a supervised classification technique where the number of categories and the sample that belong to each category are previously defined. The method supplies a number of orthogonal linear discriminant functions, equal to the number of categories minus one, that allow the samples to be classified in one or another category [21,23]. LDA was carried out using *JMP* software (version 5.01, SAS Institute Inc., Cary, NC, USA) on the PCA sample scores on components 1 and 2, which gave the highest level of separation in the PCA models developed. Statistics calculated for the calibrations included the coefficient of determination in calibration (R^2) and the root mean square of the standard error of cross validation (RMSECV).

3. Results and discussion

Fig. 1 shows the mean mass spectrum of volatiles for the Riesling and unwooded Chardonnay wines analysed. It was observed that some differences exist between the two wine varieties in the ions at *m*/*z* 55, 61, 64, 70 and 73, and greater differences were found at *m*/*z* 88, 101, 115, 127 and 147. Although these ions might be characteristic of esters and other volatile compounds

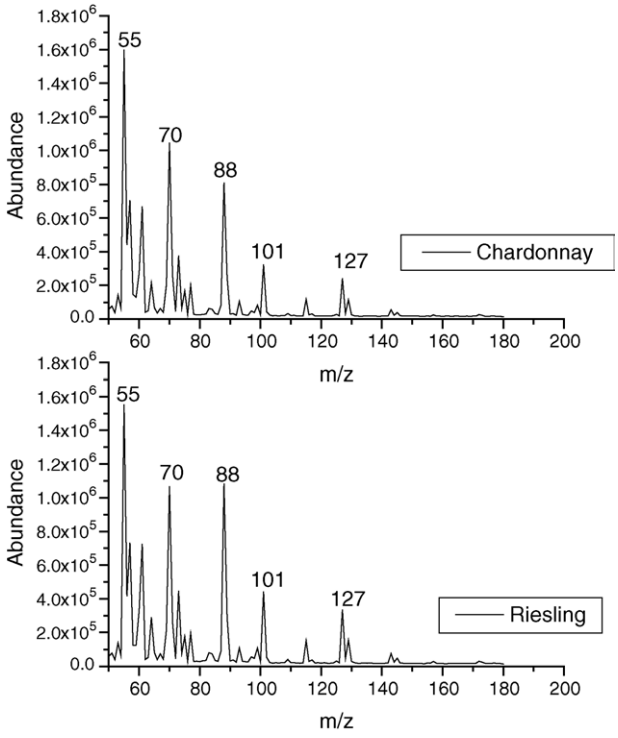

Fig. 1. Mean mass spectra of Riesling and unwooded Chardonnay wines obtained by MS_enose.

in the wine matrix [17,25,27], the identification of the ions or relating them to the chemical composition of the headspace is beyond the scope of this study.

A PCA model with three PCs explains 90% of the variation in the MS e_nose data for the Riesling and unwooded Chardonnay wines. The score plot for the first two principal components (PC1 versus PC2) is shown in Fig. 2. The score plot reveals that separation along PC1, which accounted for 62% of the variation in the sample set, while separation along PC2 accounted for 17% of the variation in the sample set. Fig. 3 shows the eigenvectors for the first three PCs indicating only the ions between *m*/*z* 100 and 180. The visual inspection of the eigenvectors confirms that the fragment ions *m*/*z* 101, 118, 127 and 147 are important variables in the differentiation of the two wine varieties. From these results and those from an additional study [17,31], it is possible that volatile compounds in the wine matrix associated with esters might explain the separation between the varieties. Similar results were reported by other authors [27].

Table 1 compares the statistics for the DPLS models before and after normalisation as pre-treatment of the MS e_nose data. The results highlighted that normalisation substantially improves the RMSECV for the calibration models, confirming the assumption that drifts or changes in the instrument might affect the calibrations when multivariate methods are applied. The results (R^2) show that more than 90% of the variations in the DPLS models were explained when normalisation was used compared with non-normalisation (less than 70%).

Table 2 compares the classification results obtained using DPLS and LDA classification methods after normalisation using all the ions (*m*/*z* 50–180). Correct classification levels of 93%

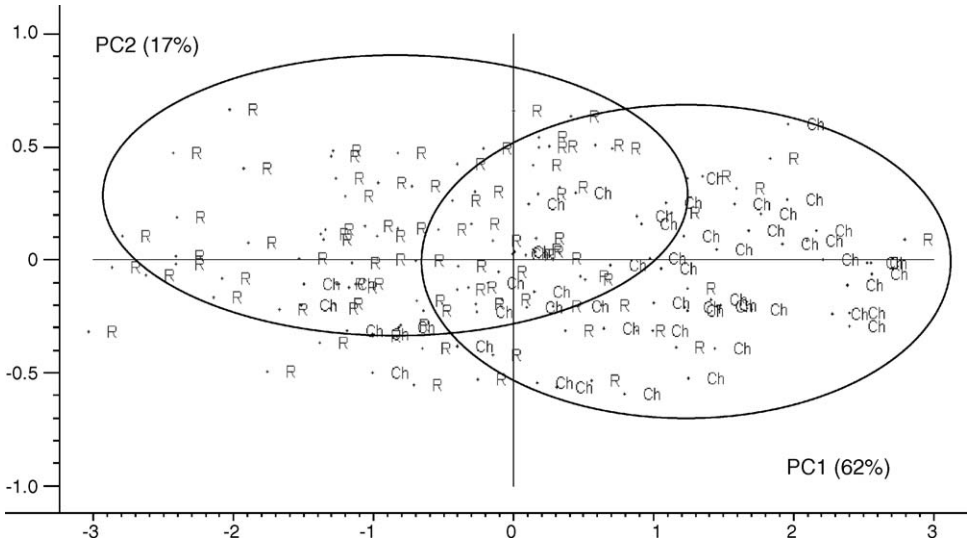

Fig. 2. Score plot of the first two principal components in MS_enose profiles of Riesling (R) and unwooded Chardonnay wines (Ch).

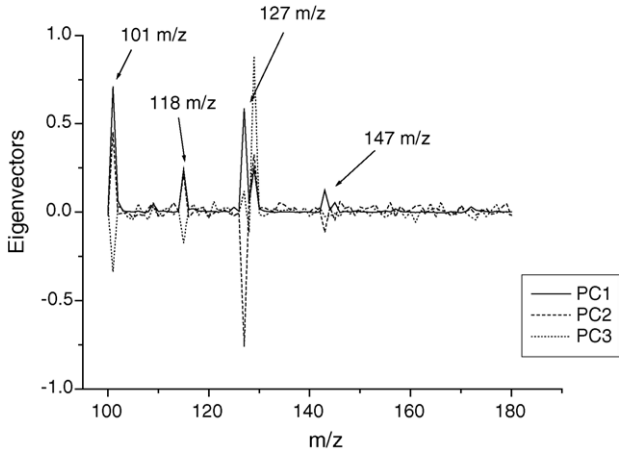

Fig. 3. Eigenvectors for the first three principal components (range *m/z* 100–180).

Table 1
Comparison of DPLS calibration statistics before and after smoothing and normalisation of the MS e_nose data ($n = 150$)

	RMSECV	R^2	Number of PLS factors
Smoothing			
m/z 50–180	0.36	0.70	3
m/z 100–180	0.40	0.40	2
Smoothing and normalization			
m/z 50–180	0.13	0.94	5
m/z 100–180	0.16	0.92	5

RMSECV: root mean square of the standard error of cross validation, R^2: coefficient of determination in calibration.

Table 2
Percentage of correct classification results for LDA and DPLS analysis of white wine varieties

	% Correct classification	
	LDA	DPLS
Riesling ($n = 75$)	82	93
Unwooded Chardonnay ($n = 75$)	73	93

were achieved by DPLS models in both wine varieties. On the other hand, LDA showed lower levels of correct classification, 73% and 82% for unwooded Chardonnay and Riesling wines, respectively. Note that only two PCs were used when LDA models were developed and this might explain the low % of correct classification obtained. It was noticed that old Riesling wines (1993, 1996 and 1997 vintages) were misclassified using both DPLS and LDA discriminant methods. As might be expected, these wines had low ester content and were more oxidised than the other Riesling wines used in this study [17,28]. Young Chardonnay wines were also misclassified, possibly due to their fruity and fresh characters, which gave similar aroma notes to those observed in typical Australian Riesling wines [26–28,31].

Fig. 4 shows the DPLS regression plot for the discrimination of both white wine varieties using MS e_nose data. Generally, there was observed a good separation of samples by variety; however, some samples did overlap. These results were consistent with a previous report using Vis-NIR, where overlap among

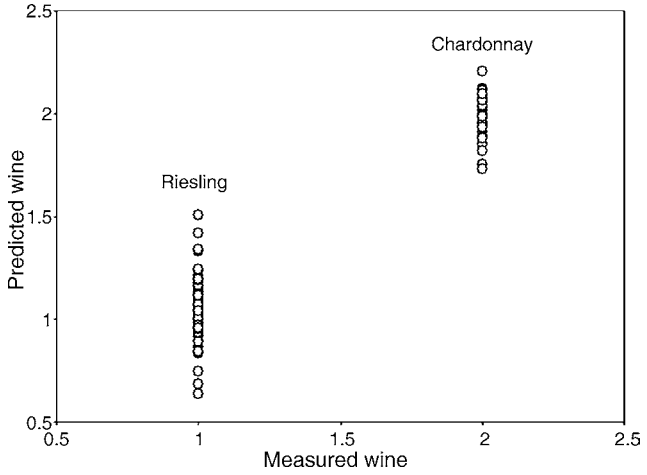

Fig. 4. Discriminant PLS (DPLS) plot of Riesling (1) and unwooded Chardonnay (2) wines.

samples was also observed [16]. It is important to note that the samples used in the current study were commercially available bottles of wine and could not be completely verified in terms of their authenticity other than by the claim made on the label [16,29]. It is therefore possible that some of the samples were not 100% of the variety as claimed, which would explain the overlap. It should be noted that according to the Australian regulations for the label to claim a single variety, the minimum content of that variety must be 85% [29]. In fact, it was verified by the winemakers that some Chardonnay samples were blended either with Riesling (up to 5% in some cases) or with other white Australian varieties (e.g., Semillon, Sauvignon Blanc). Therefore, this result suggests that discrimination between varieties is possible, and that different aroma properties or volatile compounds present in the samples were associated with either characteristics of the variety, winemaking style or yeast strain. Similar results were found by other authors using similar wines (one Riesling and one Chardonnay) analysed by MS e_nose [30].

It is well known that ethanol is the major volatile component in the wine matrix, thus some interference can be expected with the MS e_nose analysis. According to other authors, ions related with ethanol could appear around the *m*/*z* ratios 45 and 46 [5]. Previous studies using the same varieties confirmed the existence of statistically significant differences between alcohol content of the varieties for part of the sample set used in this study [16]. Therefore, the scan range chosen for the present study for the MS e_nose avoids the ethanol, as only ions with a mass/charge ratio >50 were monitored.

Wine aroma is formed by a complex mixture of many natural and processing variables, such as degree of maturity at harvest, percent of solid present in the fermented grape juice and yeast strain [26]. The use of MS e_nose can be problematic as the composition of the headspace is monitored rather than the sample itself. It is well known that the headspace concentration of a particular substance is related to the vapour pressure, to the liquid-phase concentration of the analyte and other volatile or non-volatile substances, and to temperature. Thus, the headspace composition may not be the true reflection of the composition of the sample itself [1]. This should imply that the application of the MS e_nose is specific for the current sample set and might be difficult to extrapolate to other situations.

The combination of both MS e_nose data and chemometrics methods gave acceptable discrimination between samples of two commercial white wine varieties. The present study only uses a limited number of samples and wine varieties, and therefore, caution must be exercised in extending the applicability of the technique to discriminate between other varieties until further validation work is completed. This study showed also the potential of MS e_nose to discriminate by variety between commercially available bottles of Riesling and unwooded Chardonnay wine with accuracy of up to 95% using DPLS and around 80% using LDA classification methods. Therefore, it is suggested that MS e_nose together with chemometrics, such as PCA or discriminant techniques, might be used by the wine industry for the identification of white wine varieties or their blends. The results highlighted the importance of pre-treatment of MS e_nose data before analysis and interpretation by multivariate techniques. Even though the conventional analysis based on GC–MS provides fundamental information about the volatile compounds presents in the wine, the MS e_nose method has advantages of simplicity of sample-preparation and reduced time of analysis. Hence, conventional GC–MS methodology might be considered more appropriate when detailed compositional characterisation is required, while MS e_nose might be a choice for routine monitoring or screening of many samples. It should be considered that one additional benefit of this new approach to wine analysis is that volatile profiling of samples might be often more useful than identification and quantification of individual compounds [30]. From the results obtained it can be concluded that PCA, LDA or DPLS techniques applied to MS e_nose data offer the possibility of classify Riesling and unwooded Chardonnay wines. Further studies are needed in order to improve the calibration specificity, accuracy and robustness, and to extend the discrimination to other wine varieties or blends. The work reported here is only a feasibility study and requires further development with considerably more commercial samples of different varieties before its full potential can be realised and it is ready for adoption by the wine industry.

Acknowledgments

The authors thank Dr. Allan Pollnitz and Dr. Markus Herderich for critical comments and suggestions on the manuscript. The critical input and continual encouragement from Professor P.B. Høj, former Director of the Australian Wine Research Institute, is also acknowledged. This project is supported by Australia's grapegrowers and winemakers through their investment body the Grape and Wine Research and Development Corporation, with matching funds from the Australian government, and by the Commonwealth Cooperative Research Centres Program. The work was conducted by The Australian Wine Research Institute and forms part of the research portfolio of the Cooperative Research Centre for Viticulture (CRCV).

References

[1] A.K. Deisingh, D.C. Stone, M. Thompson, Int. J. Food Sci. Tech. 39 (2004) 587–604.
[2] I. Arvantoyannis, M.N. Katsota, P. Psarra, E. Soufleros, S. Kallinthraka, Trends Food Sci. Technol. 10 (1999) 321–336.
[3] W.J. Hurst, Electronic Noses and Sensor Array Based Systems: Design and Applications, Technomic, USA, 1999, p. 340.
[4] N. Boudaoud, L. Eveleigh, J. Agric. Food Chem. 51 (2003) 1530–1533.
[5] P.N. Bartlett, J.E. Elliot, J.W. Gardner, Food Technol. 51 (1997) 44–48.
[6] M.P. Marti, R. Boque, M. Riu, O. Busto, J. Guasch, Anal. Bioanal. Chem. 376 (2003) 497–501.
[7] C. Di Natale, F.A.M. Davide, A. D'Amico, G. Sberveglieri, P. Nelli, G. Faglia, C. Perego, Sens. Actuators 24–25 (1995) 801–804.
[8] C. Di Natale, R. Paolesse, M. Burgio, E. Martinelli, G. Pennazza, A. D'Amico, Anal. Chim. Acta 513 (2004) 49–56.
[9] C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, M. Ubigli, L. Lvova, A. Rudnitskaya, Y. Vlasov, Sens. Actuators 69 (2000) 342–347.
[10] M. Penza, G. Cassano, Food Chem. 86 (2004) 283–296.
[11] S. Buratti, S. Bendetti, M. Scampicchio, E.C. Pangerod, Anal. Chim. Acta 525 (2004) 133–139.

[12] C. Pinheiro, C.M. Rodrigues, Th. Schafer, J.G. Crespo, Biotech. Bioeng. 77 (2002) 632–640.
[13] A. Guadarrama, J.A. Fernandez, M. Iñiguez, J. Souto, J.A. de Saja, Sens. Actuators 77 (2001) 401–408.
[14] W.U. Cynkar, I.L. Francis, H. McCarthy, E. Gomez, E.J. Bartowsky, in: R. Blair, P. Williams, P.B. Høj (Eds.), Proceedings of the 11th Australian Wine Industry Technical Conference, 2001, p. 257.
[15] Ch. Cordella, I. Moussa, A.-C. Martel, N. Sbirrazzuoli, L. Lizzani-Cuvelier, J. Agric. Food Chem. 50 (2002) 1751–1764.
[16] D. Cozzolino, H.E. Smyth, M. Gishen, J. Agric. Food Chem. 52 (2003) 7701–7711.
[17] H.E. Smyth, D. Cozzolino, I.L. Francis, in: S.M. Bell, K.A. de Garis, C.G. Dundon, R.P. Hamilton, S.J. Partridge, G.S. Wall (Eds.), Proceedings of The Australian Society of Viticulture and Oenology, 2003, pp. 56–58.
[18] M. Wirthensohn, I.L. Francis, R. Gawel, G.P. Jones, in: P. Iles, E. Knox (Eds.), Proceedings of the Australian International Symposium on Analytical Science, Royal Australian Chemical Institute, Melbourne, Australia, 1999, pp. 75–78.
[19] I. Marcos Lorenzo, J.L. Perez Pavon, M.E. Fernandez Laespada, C. Garcia Pinto, B. Moreno Cordero, L.R. Henriques, M.F. Peres, M.P. Simoes, P.S. Lopes, Anal. Bional. Chem. 374 (2002) 1205–1211.
[20] A. Klieber, M.N. Muchui, Aus. J. Exp. Agric. 42 (2002) 201–205.
[21] T. Naes, T. Isaksson, T. Fearn, T. Davies, A User-friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002, p. 420.
[22] M. Martens, H. Martens, Multivariate analysis of quality: an introduction, John Wiley and Sons, Chichester, U.K., 2001, p. 400.
[23] M.J. Adams, Chemometrics in Analytical Spectroscopy, Analytical spectroscopy monographs, The Royal Society of Chemistry, 1995, p. 216.
[24] M. Otto, Chemometrics, Wiley VCH, 1999, p. 314.
[25] D.H. Williams, I. Fleming, Spectroscopic Methods in Organic Chemistry, McGraw Hill, 1995, p. 329.
[26] B. Rankine, Making Good Wine: A Manual of Winemaking Practice for Australian and New Zealand, The Macmillan Company of Australia, 1998, p. 374.
[27] U. Fischer, D. Roth, M. Christmann, Riesling wines. Food Qual. Pref. 10 (1999) 281–288.
[28] P. Iland, P. Gago, Australian Wine: Styles and Tastes, Patrik Iland Wine Promotions, Australia, 2002, p. 202.
[29] The Australian Wine and Brandy Corporation, Wine law, making and labelling, 15 May 2003; Available [http://www.awbc.com.au/winelaw/wine_label_law_pdfs/WLLABL15May03.pdf] via the INTERENT. Accessed 2003 August 2003.
[30] I.L. Francis, W.U. Cynkar, Tech. Rev. 128 (2000) 1–5.
[31] H.E. Smyth, The compositional basis of the aroma of Riesling and unwooded Chardonnay wine, PhD Thesis, University of Adelaide, Australia, 2005, p. 220.