Strategies to minimise matrix-related error with near infrared analysis of wine grape quality parameters

R.G Dambergs[a], D. Cozzolino[a], W.U. Cynkar[a], M.B. Esler[b], L.J. Janik[a], I.L. Francis[a] and M. Gishen[a].

[a] *Australian Wine Research Institute. Waite Road, Urrbrae. PO Box 197, Glen Osmond, South Australia, 5064. Fax: + 61 8 8303 6601. Email: bob.dambergs@awri.com.au*

[b] *Present address: Chemistry Section, BIPM, Pavillon de Breteuil, F-92312 Sevres Cedex, France.*

Introduction

The ability to objectively measure grape quality has become an important need in the Australian wine industry[1]. To this end we have developed near infrared (NIR) spectroscopy methods for analysis of total anthocyanins (colour), total soluble solids (TSS, predominantly sugars) and pH of red wine grapes.[2,3] Our work has concentrated on the measurement of anthocyanins because although it has been known for some time that within certain specific limitations grape and wine colour generally correlates with wine quality,[4,5] there has been slow uptake of colour assessment by commercial wine producers mainly due to the relative complexity of the standard assay procedure[6].

With grapes from one season and from limited growing regions, it has been possible to produce calibrations for NIR prediction of colour with sufficient accuracy for commercial use.[7] However, when attempting to produce a universal calibration spanning many seasons, growing regions and grape varieties, the performance of partial least squares regression in prediction of colour was limited by pronounced non-linearity[8]. A similar problem also occurred with pH, but not with TSS. Selection of sample subsets based on vintage, grape variety and growing region, resulted in reduction of calibration non-linearity and increased accuracy – it was therefore concluded that matrix variation related to these classification variables was at the root of the problem[8]. The current study focussed on minimising this matrix-related error, potentially allowing application of calibrations for grape anthocyanins across the Australian wine industry, regardless of growing region, grape variety and spanning growing seasons.

Methods

Sample preparation

Grape samples were stored as frozen berries (approximately 100 g, with stems and stalks removed) and when required for analysis were thawed and homogenised while cool (5-10°C) with an *Ultra-turrax T25N* homogeniser (Janke and Kunkel GmbH Co. Germany) at 25,000 rpm for 60 seconds.

Reference analysis

To perform the reference analysis for total anthocyanins,[6] aliquots of homogenates (approximately 1 g) were scooped into pre-tared centrifuge tubes and their masses recorded. The

Near Infrared Spectroscopy: Proceedings of the 11th International Conference.

samples were then extracted in 10 mL of 50% v/v aqueous ethanol, for 1 hour, with continuous gentle agitation. The extract was clarified by centrifugation then the supernatant was diluted 1:10 with 1N HCl to allow expression of the ionised, coloured (red) form of anthocyanins (minimum 3 hour incubation). The concentration of total anthocyanins was calculated from the absorbance at 520 nm and expressed as milligrams of anthocyanin (calculated as malvidin-equivalents) per gram of fruit weight.

Sample scanning and data analysis

Homogenates were scanned without temperature equilibration, in a FOSS *NIRSystems 6500* (FOSS NIRSystems, Silver Spring, Maryland, USA), in reflectance mode at 2 nm intervals over the wavelength range of 400–2500 nm. A reference scan was performed before each sample. Spectra were stored as the average of 32 scans. Scanning control was performed with the *Vision* software package (FOSS NIRSystems, Silver Spring, Maryland, USA).

Chemometric analyses were performed with *WinISI* software (version 1.5, Infrasoft International, LLC, USA). The standard data pre-treatment was to use the 600–1800 nm segment, standard normal variate and detrend transformations, with first derivative (4 data point gap) and 4 data point smoothing. Calibrations were developed using modified partial least squares (MPLS) regression or LOCAL regression[9]. A total of 2287 samples, spanning three growing seasons, ten grape varieties and ten growing regions were used for the analysis. The sample set was sorted by reference value then every tenth sample was removed to create a separate validation set. When cross validation was performed, four validation groups were used.

Results and discussion

When using MPLS, a comparison of NIR-predicted colour with the reference value showed pronounced non-linearity, with under-prediction at extreme colour values and over-prediction at median colour values (Figure 1.a). With the LOCAL algorithm a 40% decrease in standard error of prediction (SEP) for colour was observed in comparison with MPLS calibrations. LOCAL calibrations were linear, although prediction error did appear to increase at higher analyte values (Figure 1.b).

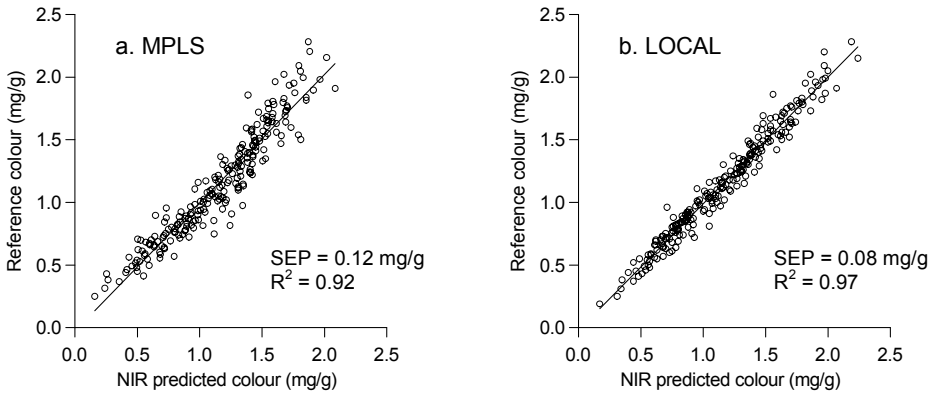

Figure 1. Comparison of calibration performance of (a) MPLS and (b) LOCAL, using a 10% subset of the main sample set as a validation set.

It was initially assumed that the improvements with the LOCAL method were due to the selection of samples with spectral similarity[9] (i.e. selection of samples based on the "x" variable). Close examination of the LOCAL sets (100 samples per set) revealed that the average analyte value closely correlated with the analyte value of the prediction sample (i.e. selection of LOCAL sets also resulted in matching of the "y" variable). Based on this information another calibration strategy was tested: the database was manually sub-divided on the basis of colour values. The "mid-range" samples had colour values of 0.5–1.5 mg/g, a range suggested to be critical for quality discrimination of grapes from warm-climate growing regions[7]. The "extreme-range" samples had colour values <0.5 mg/g or >1.5 mg/g. Mid-range MPLS calibrations had an SECV comparable to the SEP for full range LOCAL calibrations, but poorly predicted the extreme-range samples (Table 1), albeit with a high R^2 that was probably an artefact of the large analyte range. A simple calibration strategy, without using LOCAL, may be to create separate MPLS calibrations to suit the analyte range rather than attempting to incorporate all samples in the one calibration.

Table 1. Comparison of MPLS calibration statistics for NIR prediction of colour using calibration sets of all samples with those using subsets selected by colour value. The standard error of cross validation (SECV) and standard error of prediction (SEP) are expressed as milligrams anthocyanin per gram of fruit. N is the number of samples.

Calibration set	N	Validation set	SECV	SEP	R^2
All samples	2287	-	0.13	-	0.92
mid-range colour	1618	-	0.09	-	0.89
mid-range colour	1618	extreme-range colour	-	0.16	0.92

Conclusion

An important finding of this study was that selection of calibration sets with the LOCAL algorithm, which is done on the basis of spectral similarity, coincidentally resulted in selection of samples with similar analyte concentrations i.e. with this particular analyte/matrix combination it effectively operated as "LOCALxy" and gave superior prediction performance. Examination of the LOCAL sets also showed that vintage, variety and growing region were not necessarily barriers to spectral matching, provided that the database contained samples with analyte values similar to the unknown sample. The implication of this observation is that an effective spectral database for measurement of total anthocyanins in red grapes may not necessarily require samples from all growing regions, and varieties, but must be representative of the full expected range of analyte values and with respect to this, the database may require careful monitoring with regard to seasonal variations.

The effect of colour concentration range on NIR calibration accuracy may be related to sample matrix differences correlated with colour concentration, for example samples with high colour values may contain a proportion of shrivelled, partially dehydrated fruit. Also the reference colour analysis method does not discriminate between different types of anthocyanins, which may vary in relative concentration with colour range. Finally there may be effects on the reference method, such as incomplete sample extraction, at extreme colour values. Further investigation is required.

Acknowledgements

This work has involved a collaboration of research and industry personnel from The Australian Wine Research Institute, Hardy Wines (especially Audrey Lim, Chris Bevin and Alex Sas),

McGuigan-Simeon Wines (especially Bruce Kambouris), Orlando Wyndham (especially Inca Lee and Russell Johnstone), and Southcorp/Rosemount (especially Peter Hayes and Mark Smith) and forms part of the research portfolio of the Cooperative Research Centre for Viticulture. The authors acknowledge the contributions of their many colleagues at the Australian Wine Research Institute, especially Elizabeth Waters, for ongoing support. The critical input and continual encouragement from Professor P.B. Høj, Director of the Australian Wine Research Institute, is also acknowledged. The financial support of Australia's grape growers and winemakers through their investment body the Grape and Wine Research and Development Corporation, with matching funds from the Federal Government, and by the Commonwealth Cooperative Research Centres Program is acknowledged.

References

1. M. Gishen, P.G. Iland, R.G. Dambergs, M.B. Esler, I.L. Francis, A. Kambouris, R.S. Johnstone, and P.B. Høj, in *Proceedings of the eleventh Australian wine industry technical conference*, Ed by R.J. Blair, P. Williams and P.B Høj, Winetitles, Adelaide, South Australia, p. 188–194 (2002).
2. M. Gishen and R.G. Dambergs, *Australian Grapegrower and Winemaker, Annual Technical Issue,* No. 414a: p. 43–45, 47 (1998).
3. R.G. Dambergs, A. Kambouris, M. Gishen, I.L. Francis, in *Modern Viticulture- meeting market expectations, Proceedings of the Australian Society of Viticulture and Oenology,* Winetitles, Adelaide, South Australia, p. 45–47 (1999).
4. T.C. Somers and M.E. Evans, *J. Sci. Food Agric.* **25**: 1369–1379 (1974).
5. I.L. Francis, P.G. Iland, W.U. Cynkar, M. Kwiatowski P.J. Williams, H. Armstrong, D.G. Botting, R. Gawel and C. Ryan, in *Proceedings of the tenth Australian wine industry technical conference*, Ed by R.J. Blair, P. Williams and P.B Høj, Winetitles, Adelaide, South Australia, p. 104–108 (1999).
6. P.G. Iland, W. Cynkar, I.L. Francis, P.J. Williams, and B.G. Coombe, *Australian Journal of Grape and Wine Research,* **2:** 171–178 (1996).
7. A.M. Kennedy, in *Proceedings of the eleventh Australian wine industry technical conference,* Ed by R.J. Blair, P. Williams and P.B Høj, Winetitles, Adelaide, South Australia, p. 199–205 (2002).
8. M.B. Esler, M. Gishen, I.L. Francis, R.G. Dambergs, A. Kambouris, W.U. Cynkar and D.R. Boehm, in *Near Infrared Spectroscopy: Proceedings of the 10th International Conference*, Ed by R.K. Cho and A.M.C. Davies, NIR Publications, Chichester, UK, p. 249–253 (2002).
9. J.S. Shenk, M.O. Westerhaus and P. Berzaghi, *J. Near Infrared Spectrosc.* **5**: 223–232 (1997).