Differentiation Between Wines According to Geographical Regions in the Western Cape (South Africa) Using Multivariate Analyses Based on Selected Chemical Parameters in Young Red Wines

P. P. Minnaar[1]* and M Booyse[2]

(1) ARC Infruitec-Nietvoorbij, Private bag X5026, 7599 Stellenbosch, South Africa
(2) ARC Biometry Unit, Private Bag X5013, 7599 Stellenbosch, South Africa

Submitted for publication: July 2004
Accepted for publication: October 2004
Key words: Pinotage, Shiraz, Cabernet Sauvignon, chemical parameters, multivariate analysis

The aim of this study was to differentiate between wines of three diverse wine regions, namely the districts Paarl, Stellenbosch and Swartland, in the Western Cape on the basis of chemical parameters, namely pH, volatile acid, total acids, malic acid, lactic acid, glucose, fructose, ethanol and glycerol. A total of 140 young red wines (Pinotage, Cabernet Sauvignon and Merlot) obtained from the 2003 National Young Wine Show were analysed. Classification methods such as stepwise discriminant analysis, canonical discriminant analysis and linear discriminant analysis were applied to the data sets. A total classification of 34.04% for Cabernet Sauvignon, 42.55% for Pinotage and 31.25% for Shiraz was determined. It was concluded that, under the conditions of this study, separation of the wines according to geographical regions was unsuccessful using the selected parameters.

Differentiation of wines between regions is important for the global wine trade, but it presents a difficult problem that could be solved by quantification of a variety of chemical constituents. Geographical differentiation of wines is possible by applying pattern recognition techniques to data obtained from a variety of analytical techniques. Several studies have been performed to differentiate wines according to their geographical origin. A number of authors have attempted wine differentiation based on chemical parameters. Differentiation of wine from France and Germany was achieved on the basis of ICP and GC data of trace elements, phenols and classic enological parameters (Kwan & Kowalski, 1978; Etievant & Schlich, 1988). The isotope ratios $(D/H)_I$ and $(D/H)_{II}$ determined by means of SNIF-NMR and trace elements determined by ICP-MS, were measured in wine originating from Bordeaux to characterise the geographic origin (Martin *et al.,* 1999). Day *et al.* (1995) determined $(D/H)\frac{Q}{W}$, $(D/H)_I$, $(D/H)_{II}$, $\delta^{18}O\frac{Q}{W}$ and, $\delta^{13}C\frac{Q}{A}$ ratios, including elemental composition of wine from Burgundy, using 2H-MNR and AA to determine geographical origin. Authenticity and geographical origin of wines from Slovenia were investigated measuring $^{13}C/^{12}C$ and $(D/H)_I$ ratios by means of IRMS and SNIF-NMR (Ogrinc *et al.,* 2001). Differentiation between German, Spanish, Italian and French wines was achieved by using stepwise discriminant analysis, canonical discriminant analysis and principal component analysis applied to amino acids, trace elements, phenols and classic enological parameters (Sivertsen *et al., 1999*; Kallithraka *et al.*, 2001; Pérez-Magariño *et al.*, 2002; Brescia *et al.*, 2003). The possibility of using the relationship between certain phenolic compounds as a tool for geographical discrimination and authenticity determinations of South African red wines was investigated by Rossouw & Marais (2004).

The aim of this study was a preliminary investigation to differentiate between young red wines from five distinct wine regions/districts in the Western Cape (South Africa). Multivariate analyses based on specific classic enological parameters, obtained by rapid analysis on the GrapeScan instrument, were used.

MATERIALS AND METHODS

Wine origin

The investigation included 140 commercial young red wines representing Pinotage, Shiraz and Cabernet Sauvignon (Table 1). These wines were collected from entries in the 2003 National Young Wine Show. All the wines came from the 2003 vintage and contained 100% of the specific cultivar.

Wine analysis

Wine samples (50 mL) were collected directly from sealed bottles. The samples were submitted for analysis without sample preparation. A Grapescan (FT 120, supplied by Foss Electric, Denmark), calibrated for South African wines, was utilised to quantify pH, ethanol, volatile acid, total acids, glucose, fructose, malic acid, lactic acid and glycerol. At the time of analysis these parameters were the only parameters for which the instrument had been calibrated under local conditions.

Statistical analysis

Univariate procedures were used to determine normality of the residuals for each variable. Scatter plots of the variables were drawn to determine multivariate normality. The assumption of normality was confirmed. For this reason, linear discriminant analysis (LDA), stepwise discriminant analysis, to select a subset of quantitative variables and canonical variate plots (CVA plots)

*Corresponding author: E-mail address: minnaarP@arc.agric.za

Acknowledgements: The authors sincerely thank the South African Wine Industry (Winetech) and the Agricultural Research Council for financial support. The donation of the wines by the South African National Wine Show Association is appreciated.

TABLE 1

Grape cultivar, wine region, district and number of wines (n).

Pinotage		Shiraz		Cabernet Sauvignon	
Region/District	**n**	**Region/District**	**n**	**Region/District**	**n**
Paarl	13	Paarl	12	Paarl	15
Stellenbosch	10	Stellenbosch	13	Stellenbosch	15
Swartland	6	Swartland	6	Swartland	6
Olifants River Valley	6	Olifants River Valley	6	Olifants River Valley	1
Breede River Valley	10	Breede River Valley	11	Breede River Valley	10
Total	**45**		**48**		**47**

were used to differentiate between wines according to geographical regions and districts. All statistical analyses were conducted using SAS version 8.2 (SAS, 1999).

RESULTS AND DISCUSSION

Stepwise discriminant analysis was performed on the full set of variables for each grape cultivar to obtain the R^2, P and F values (Table 2).

Canonical discriminant analysis

Application and results of canonical discriminant analysis using the full set of nine variables measured in Pinotage, Cabernet Sauvignon and Shiraz are represented in Figures 1, 2 and 3. These results indicate that natural groupings are not formed when all nine variables are considered simultaneously.

For this reason it was decided to select the best subset of mea-

TABLE 2

The full set of variables, P values, F values and R^2 values for each grape cultivar obtained from stepwise discriminant analysis, for use in canonical discriminant analysis and linear discriminant analysis.

Cultivar	Discriminating variables	R^2	F-value**	P-Value
Cabernet Sauvignon	**Ethanol***	**0.2712**	**3.91**	**0.0087**
	Fructose	0.0566	0.62	0.6540
	Glucose	0.0742	0.82	0.5187
	Glycerol	0.0399	0.43	0.7887
	Lactic Acid	0.1077	1.24	0.3103
	Malic acid	0.0242	0.25	0.9053
	pH	**0.2647**	**3.78**	**0.0103**
	Total acids	**0.1727**	**2.19**	**0.0863**
	Volatile acid	0.0416	0.44	0.7756
Pinotage	**Ethanol**	**0.1461**	**1.40**	**0.2435**
	Fructose	**0.1846**	**1.86**	**0.1233**
	Glucose	**0.1184**	**1.10**	**0.3747**
	Glycerol	**0.1236**	**1.16**	**0.3467**
	Lactic Acid	**0.1482**	**1.43**	**0.2351**
	Malic acid	**0.1444**	**1.38**	**0.2504**
	pH	0.0947	0.86	0.5176
	Total acids	0.0568	0.49	0.7789
	Volatile acid	0.0617	0.54	0.7451
Shiraz	**Ethanol**	**0.2101**	**2.86**	**0.0346**
	Fructose	0.0613	**0.69**	0.6062
	Glucose	**0.1503**	**1.90**	**0.1275**
	Glycerol	**0.1006**	**1.20**	**0.3236**
	Lactic Acid	0.0072	0.08	0.9892
	Malic acid	**0.1363**	**1.70**	**0.1684**
	pH	0.0544	0.60	0.6615
	Total acids	0.0237	0.25	0.9054
	Volatile acid	0.0944	1.10	0.3715

R^2 = R squared; P = Probability of the F-value; F = Statistics of decision-making criteria.
* = The most powerful discriminant variables are shown in bold; ** = F-values not significant.

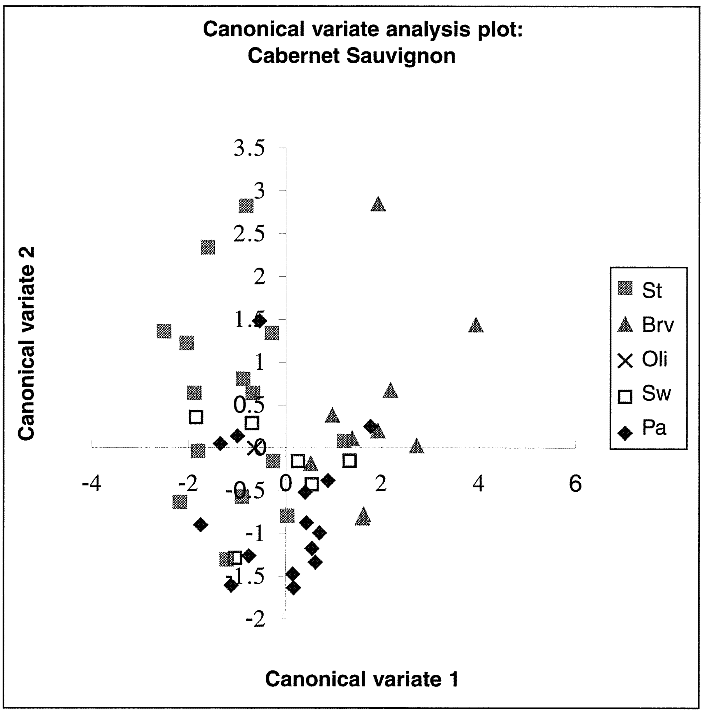

FIGURE 1

Plots of the first two canonical variables for Cabernet Sauvignon, defined by nine variables [pH, ethanol, volatile acid, total acids, glucose, fructose, malic acid, lactic acid and glycerol] (Pa = Paarl; St = Stellenbosch; Sw = Swartland; Oli = Olifants River Valley and Brv = Breede River Valley).

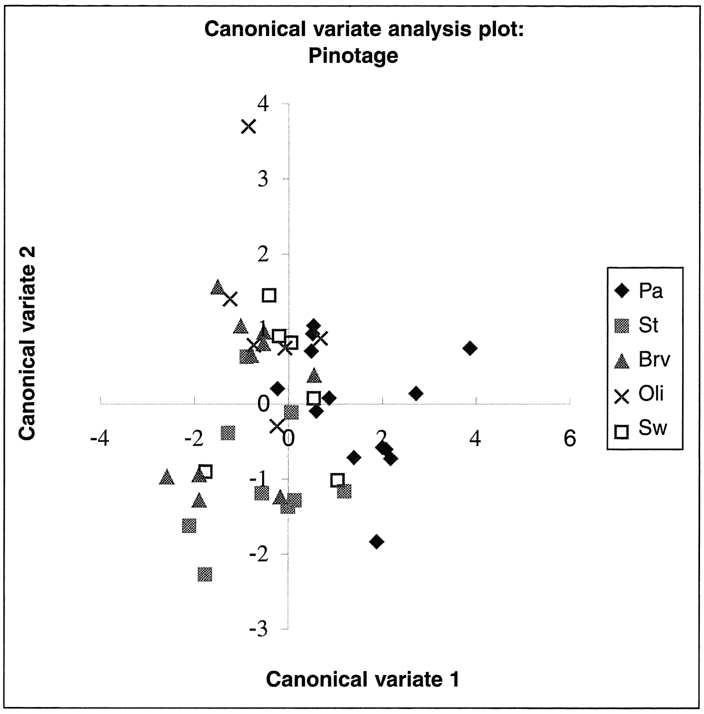

FIGURE 2

Plots of the first two canonical variables for Pinotage, defined by nine variables [pH, ethanol, volatile acid, total acids, glucose, fructose, malic acid, lactic acid and glycerol] (Pa = Paarl; St = Stellenbosch; Sw = Swartland; Oli = Olifants River Valley and Brv = Breede River Valley).

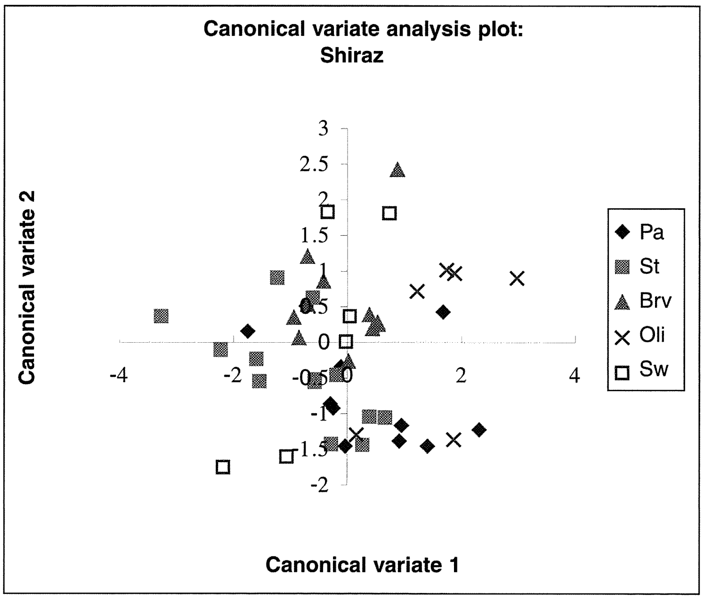

FIGURE 3

Plots of the first two canonical variables for Shiraz, defined by nine variables [pH, ethanol, volatile acid, total acids, glucose, fructose, malic acid, lactic acid and glycerol] (Pa = Paarl; St = Stellenbosch; Sw = Swartland; Oli = Olifants River Valley and Brv = Breede River Valley).

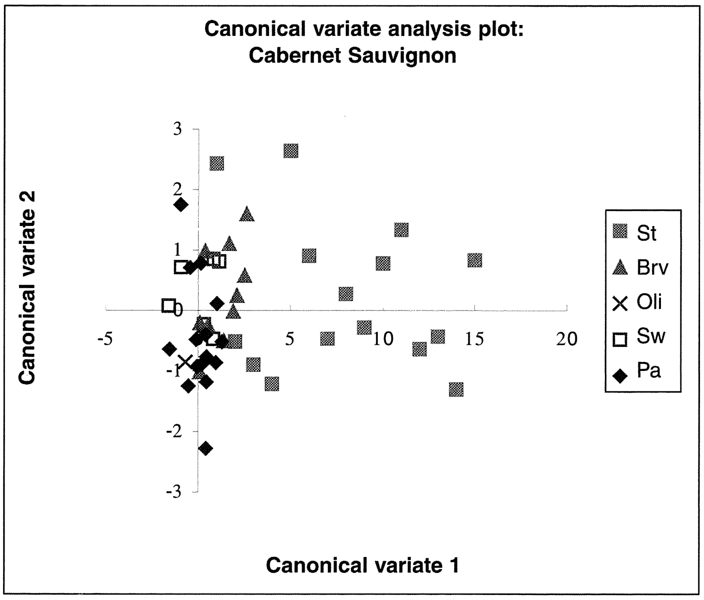

FIGURE 4

Plots of the first two canonical variables for Cabernet Sauvignon, defined by three variables [pH, total acids and ethanol] (Pa = Paarl; St = Stellenbosch; Brv = Breede River Valley; Oli = Olifants River Valley; Sw = Swartland).

surements from the nine variables using stepwise discriminant analysis. According to Klecka (1980), stepwise discriminant analysis may be performed to select the most powerful discriminators. The criteria for inclusion in the subset were the P-values and F-values. According to Krzanowski (1989), the F-values could be used as indicators for inclusion in the subset, even though the probabilities of the F-values are not significant. Krzanowski (1989) suggests using the largest F-values for inclusion. In Cabernet Sauvignon, Pinotage and Shiraz, three, six and four variables respectively exhibited the most powerful discriminant abilities (Table 2). These selected variables for all three cultivars were again subjected to canonical discriminant analysis and

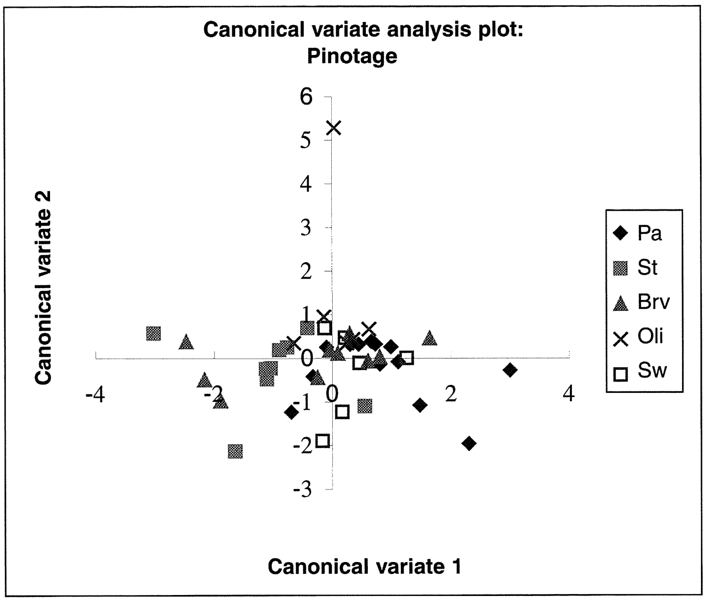

FIGURE 5

Plots of the first two canonical variables for Pinotage, defined by six variables [malic acid, lactic acid, glucose, fructose, ethanol and glycerol] (Pa = Paarl; St = Stellenbosch; Brv = Breede River Valley; Oli = Olifants River Valley; Sw = Swartland).

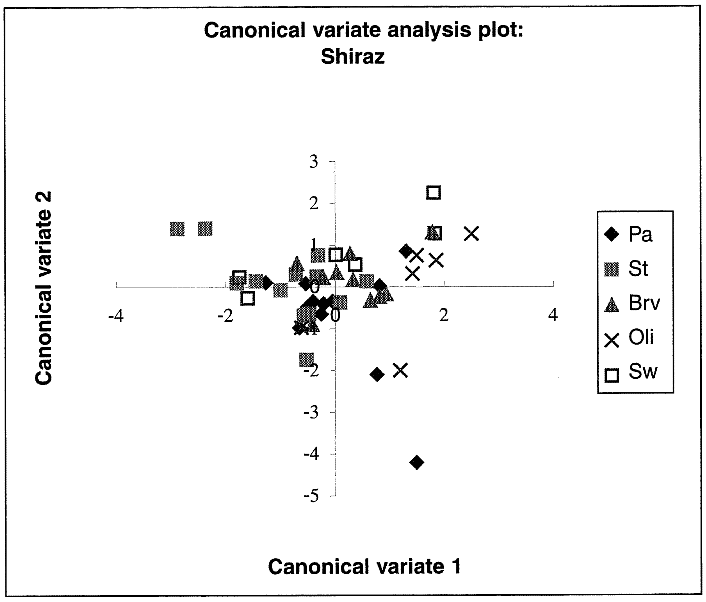

FIGURE 6

Plots of the first two canonical variables for Shiraz, defined by four variables [malic acid, glucose, ethanol and glycerol] (Pa = Paarl; St = Stellenbosch; Brv = Breede River Valley; Oli = Olifants River Valley; Sw = Swartland).

TABLE 3

Percentage correctly classified regions and districts for Cabernet Sauvignon using linear discriminant analysis.

Region/District (n = number of wines)	Percentage (%) correctly classified using cross-validation results
	Linear discriminant analysis
Breede River Valley (n = 10)	60.0%
Olifants River Valley (n = 1)	0.0%
Paarl (n = 15)	46.6%
Stellenbosch (n = 15)	20.0%
Swartland (n = 6)	0.0%
Total classification	**34.04%**

TABLE 4

Percentage correctly classified regions and districts for Pinotage using linear discriminant analysis.

Region/District (n = number of wines)	Percentage (%) correctly classified using cross-validation results
	Linear discriminant analysis
Breede River Valley (n = 10)	20.00%
Olifants River Valley (n = 6)	0.00%
Paarl (n = 13)	38.46%
Stellenbosch (n = 9)	55.56%
Swartland (n = 6)	0.00%
Total classification	**42.55%**

TABLE 5

Percentage correctly classified regions and districts for Shiraz using linear discriminant analysis.

Region/District (n = number of wines)	Percentage (%) correctly classified using cross-validation results
	Linear discriminant analysis
Breede River Valley (n = 11)	36.3%
Olifants River Valley (n = 6)	16.6%
Paarl (n = 12)	33.3%
Stellenbosch (n = 13)	46.1%
Swartland (n = 6)	0.0%
Total classification	**31.25%**

linear discriminant analysis, to determine whether classification among geographical regions and districts could be established. These two discriminant methods are similar, but use different approaches. The groups were defined according to geographical origins. Quantitative estimation of the classification power of the calculated discrimination model is given for each cultivar in a classification matrix. Application and results of canonical discriminant analysis using the selected subset of variables are presented in Figures 4, 5 and 6.

These results clearly indicate that general geographic discrimination was again unsuccessful; however, Cabernet Sauvignon from Stellenbosch could to some degree be separated from similar wines in other regions and districts (Fig. 4).

Linear discriminant analysis

Results of linear discriminant analysis, using discriminating variables obtained from stepwise discriminant analysis, for Cabernet

Sauvignon, Pinotage and Shiraz, are listed in Tables 3, 4 and 5, respectively. Quantitative estimations of the classification power of the calculated discrimination model were used. Cross-validation (using linear discriminant analysis), determined the percentage of correctly classified regions.

The total classification of 34.04% for Cabernet Sauvignon, 42.55% for Pinotage and 21.25% for Shiraz was determined. Separation of the geographical regions and districts using linear discriminant analysis was therefore unsuccessful, using the selected variables in Table 2.

CONCLUSIONS

These results have shown that, under the prevalent conditions at the time of the study, pH, ethanol, volatile acid, total acids, glucose, fructose, malic acid, lactic acid and glycerol cannot be used as discriminatory variables to differentiate wines according to the geographical areas in which they were produced. However, as additional parameters are established, the GrapeScan may become a valuable tool to facilitate differentiation between wines of different origins. This may contribute to authenticity studies on wine.

LITERATURE CITED

Brescia, M. A., Kosir, I. J., Caldarola, J. K. & Sacco, A., 2003. Chemometric classification of Apulian and Slovenian wines using ^{1}H NMR and ICP-OES together with HPICE data. J. Agric. Food Chem. 51, 21-26.

Day, M.P., Zhang, B. & Martin, G.J., 1995. Determination of the geographical origin of wine, using joint analysis of elemental and isotopic composition. II. Differentiation of the principal production zones in France for the 1990 vintage. J. Sci. Food Agric. 67, 113 - 123.

Etiévant, P. & Schlich, P., 1988. Varietal and geographic classification of French red wines in terms of elements, amino acids, and aromatic alcohols. J. Sci. Food Agric. 45, 25 - 41.

Kallithraka, S., Arvanitoyannis, I. S., Kefalas, P., El-Zajouli, A., Soufleros, E. & Psarra, E., 2001. Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin. Food Chem., 73, 501-514.

Klecka, W. R., 1980. Discriminant analysis. Sage University Papers. University of Minnesota.

Krzanowski, W. J., 1989. Principles of Multivariate Analysis. A User's Perspective. Clarendon Press. Oxford.

Kwan, W.O. & Kowalski, B.R., 1978. Classification of wines by applying pattern recognition to chemical composition data. J. Food Sci. 43, 1320 -1323.

Martín, G. J., Mazure, M., Jouitteau Martin, Y- L., Anguile, L. & Allain, P., 1999. Characterization of the geographic origin of Boudreaux wines by a combined use of isotopic and trace element measurements. Am. J. Enol. Vitic. 50, 409-417.

Ogrinc, N., Košir, I. J., Kocjančič, M. & Kidrič, J., 2001. Determination of Authenticity, Regional Origin and Vinatge of Slovenian Wines Using a Combination of IRMS and SNIF-NMR. J. Agric. Food Chem. 49, 1432-1440.

Pérez-Magariño, S., Ortega-Hea, M. & Gonzáles-San José, M. L., 2002. Multivariate classification of rosé wine from different Spanish protected designations of origin. Anal. Chim. Acta, 458, 187-190.

Rossouw, M. & Marais, J., 2004. The phenolic composition of South African Pinotage, Shiraz and Cabernet Sauvignon wines. S. Afr. J. Enol. Vitic. 25, 94-104.

SAS Institute, Inc., 1999, SAS/STAT User's Guide, Version 8, 1st printing, Volume 1. SAS Institute Inc, SAS Campus Drive, Cary, North Carolina 27513.

Sivertsen, H. K., Holen, B., Nicolaysen, F. & Risvik, E., 1999. Classification of French wines according to their geographical origin by the use of multivariate analyses. J. Sci. Food Agric. 79, 107-115.