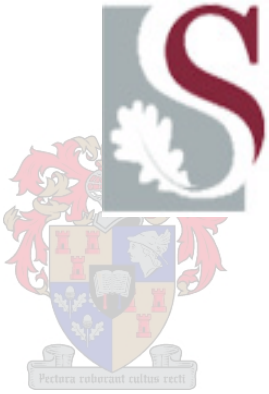
Discriminating wine yeast strains and their fermented wines: an integrated approach.

by

Charles D. Osborne



Thesis presented in partial fulfilment of the requirements for the degree of Master in Science at Stellenbosch University

Supervisor:
Prof P van Rensburg

Co-supervisors:
Dr HH Nieuwoudt
Prof KH Esbensen

December 2007

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously, in its entirety or in part, submitted it at any university for a degree.

_____ _____

Charles D Osborne **Date**

SUMMARY

The discrimination between wine yeast strains as well as between their fermented wines has been investigated in this pilot study. The study was divided in two parts, the first to investigate the discrimination between wines fermented with five different *Saccharomyces cerevisiae* yeast strains, analysed by gas chromatography (GC) and Fourier transform infrared spectroscopy (FTIR) and the second part to investigate discrimination between wine yeast strains in different liquid media and in dried form using FTIR in transmission and attenuated total reflectance (ATR) modes.

Wines from three cultivars (Clairette Blanche, Pinotage and Cabernet Sauvignon) that were fermented by five *Saccharomyces cerevisiae* yeast strains (VIN13, WE372, VIN13-EXS, VIN13-PPK and ML01) were analysed by GC and FTIR. This analysis was done on individual sample sets that consisted of the wines of each of the mentioned cultivars and also on samples drawn throughout the ageing process of these wines. The data obtained were analysed by PLS-Discrimination (PLS-discrim), a chemometric method. Using the data from both the analytical methods, discrimination was observed between wines fermented with different yeast strains in each of the two vintages (2005 and 2006) for all the cultivars. When combining the data from the two vintages no discrimination could be observed between the fermented wines. The discrimination of the fermented wines was found to be similar when using data from GC and FTIR, respectively. Since analysis with FTIR is considerably faster than analysis by GC, it would be recommended that FTIR is used for future studies of similar nature. Combining the samples into one set consisting of wines fermented with commercial wine yeast strains and wines fermented from closely related wine yeast strains (the parental strain and two genetically modified versions thereof (VIN13, VIN13-EXS and VIN13-PPK), those fermented with closely related stains did not show good discrimination from each other. Discrimination was found between wines fermented with genetically modified (GM) wine yeast strains and those fermented with non-GM wine yeast strains. This was done on a limited number of yeast strains and a larger study is needed to confirm these results. As this is the first study of this nature and differences seen could be as result of the different phenotypes.

It was shown that it is possible to use both FTIR-transmission and FTIR-ATR (attenuated total reflectance) to discriminate between different wine yeast strain phenotypes. It was shown that when using FTIR-transmission there is discrimination between yeast samples suspended in yeast-peptone-dextrose (YPD) and in water. Dried yeast samples could be discriminated when the yeast samples were in a granular,

powder form or in a pellet form, using FTIR-ATR. It was possible to discriminate between the closely related yeast strain phenotypes using FTIR-ATR.

In this pilot study it was shown that there can be discriminated between different wine yeast strains and also between the wines fermented with different wine yeast strains. It is recommended that further studies be conducted to refine and expand the study.

OPSOMMING

In hierdie loods studie is die onderskeiding van wyngiste en hul gegisde wyne getoets. Die studie is verdeel in twee ondersoeke, die eerste deel handel oor die onderskeiding tussen wyne wat gegis is met vyf verskillende *Saccharomyces cerevisiae* wyngiste wat geanaliseer is met gaschromatografie (GC) en Fourier transform infrarooi (FTIR) spektroskopie en die tweede handel oor die onderskeiding van wyngiste in gedroogde vorm en in verskillende vloeistof media wat deur midddel van FTIR in transmissie funksie en in verswakte totale weerkaatsing (ATR) funksie.

Wyne van drie kultivars (Clairette Blanche, Pinotage en Cabernet Sauvignon) wat gegis is deur vyf *Saccharomyces cerevisiae* gisrasse (VIN13, WE372, VIN13-EXS, VIN13-PPK en ML01) is geanaliseer met behulp van GC en Fourier transform (FTIR). 'n Kemometriese tegniek, Parsiële kleinste kwadrate discriminant analiese (PLS-Discrim), is gebruik om die data te analiseer. Deur gebruik te maak van die data van GC en FTIR is daar onderskeiding gevind tussen die wyne wat gegis is met die verskillende wyngiste vir elk van die twee oesjare (2005 en 2006) vir al die kultivars. Daar is egter geen onderskeiding gevind tussen die wyne nadat die data van die twee oesjare saamgevoeg is nie. Deur gebruik te maak van onderskeidelik GC en FTIR data, is daar in die kemometriese analiese geen verskil gevind in hul vermoë om tussen die wyne te onderskei nie. Verder is die analiese met behulp van FTIR aansienlik vinniger as met GC en dit word voorgestel dat die analiese in toekomstige studies met FTIR gedoen word. Die samevoeging van wyne wat gegis is met kommersiële gisrasse en na-verwante gisrasse (dit is gisrasse wat geneties gemanipuleer (GM) is van dieselfde oorspronklike gisras en die oorspronklike gisras (VIN13-EXS, VIN13-PPK en VIN13)) in dieselfde datastel, het tot gevolg gehad dat daar nie goeie onderskeiding was tussen die die wyne van na-verwante gisrasse nie. Dit is verder gevind dat daar onderskeiding was tussen wyne wat gegis is met GM en nie-GM wyngiste. Aangesien hierdie die eerste studie van sy soort is en die gistings gedoen is met betreklik min giste kan die onderskeiding moontlik wees as gevolg van verskillende phenotipes. Dit is dus noodsaaklik om verdere studies te onderneem om die resultate te bevestig.

Daar is ook bevestig dat dit moontlik is om FTIR in transmissie en in ATR funksie te gebruik om tussen verskillende wyngis fenotipes te onderskei. Onderskeiding is gevind tussen wyngisrasse wat in gis-peptoon-dekstrose en water gesuspendeer is en geanaliseer is deur FTIR in transmissie funksie. Deur gebruik te maak van FTIR-ATR is daar onderskeiding gevind tussen verskillende gedroogte giste in korrel, verpoeierde en

tablet vorm. Dit was moontlik om onderskeid te tref tussen na-verwante gisras fenotipes.

Aangesien hierdie loods studie gewys het dat daar wel onderskeid getref kan word tussen verskillende wyngiste en wyne wat gegis is met verskillende wyniste, moet verdere ondersoeke gedoen word om die studie te verfyn en te vergroot.

This thesis is dedicated to my late father for his inspiration.
Hierdie tesis is opgedra aan my oorlede vader vir sy inspirasie.

BIOGRAPHICAL SKETCH

Charles Osborne was born in Somerset West on 28 May 1971. He attended Lochnerhof Primary School and matriculated at Strand High, Strand in 1989. Charles enrolled at Stellenbosch University in 1990 and obtained a B.Eng (Chemical Engineering) in 1994. After traveling and working in industry for nine years, Charles enrolled as an MSc student at the Institute for Wine Biotechnology.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

PROF PIERRE VAN RENSBURG, Institute of Wine Biotechnology, Stellenbosch University, (supervisor), for the support and assistance;

DR HÉLÉNE NIEUWOUDT, Institute of Wine Biotechnology, Stellenbosch University, (co-supervisor), for her support, guidance, discussions and friendly smile;

PROF KIM H. ESBENSEN, ACABS, Aalborg University Esbjerg, Denmark, (co-supervisor), for all the support, guidance and knowledge;

CAMPBELL for coffee and chats, **VISHIST** and **MIA** for helping me crush early in the morning, **DANIEL** and **DANIE** for lots of laughter and all the other staff and students of the IWBT for the fun times we had;

EDMUND, **JUANITA** and **ANDY** for all the help in the experimental cellar;

KAROLIEN and **HUGH** of the Chemical Analytical Lab, for your help with the FOSS and GC;

ROGER WARREN, WarrenChem, for obtaining a sample of the commercial GM wine yeast ML01;

THE NATIONAL RESEARCH FOUNDATION AND IWBT, for financial support;

MY FAMILY for their love, support and interest in my work;

THE STEYN, RAJPAUL and **MAART FAMILIES**, for their support and interest; and

CHARLENE, without whom this venture of mine would not have been possible, for all her support, understanding and love.

PREFACE

This thesis is presented as a compilation of five chapters. Each chapter is introduced separately and is written according to the style of the journal *South African Journal of Enology and Viticulture* to which Chapter 3 and Chapter 4 will be submitted for publication.

Chapter 1 **General Introduction and Project Aims**

Chapter 2 **Literature Review**
The use of chemometrics in oenology

Chapter 3 **Research Results**
Discrimination between wines fermented by different yeast strains: a feasibility study comparing mid infrared spectroscopy with gas chromatography

Chapter 4 **Research Results**
The use of Fourier transform infrared (FTIR) spectroscopy for yeast strain phenotype discrimination

Chapter 5 **General Discussion and Conclusions**

CONTENTS

Chapter 1

INTRODUCTION AND PROJECT AIMS

1.1 INTRODUCTION

Due to increasing globalisation, the food and beverages that we consume can come from anywhere in the world. So it is understandable that consumers want information about the products they consume. This can be related to nutritional value, origin, ingredients, possible allergens or other food related issues. Many producers share this concern and strive to deliver food conforming to high quality standards and protect their products against producers misrepresenting food for economic benefit. Around the world this has led to an ongoing process of introducing legislation to protect food quality as reviewed by Reid *et al.* (2006).

In order to provide the necessary information and comply with legislation many analytical techniques have been formulated to deal with issues regarding the authentication of food products and the prediction of quality related parameters. Techniques used include spectroscopy (UV, NIR, MIR, visible, Raman), isotopic analysis, chromatography, electronic nose, polymerase chain reaction, enzyme-linked immunosorbent assay and thermal analysis (Reid *et al.*, 2006).

Modern instrumentation generates huge amounts of data for analysed samples that need to be interpreted for authentication or prediction purposes. The amounts of data that need processing led to the introduction of chemometrics with the improvements in computing capacity, as early as the 1970's, when chemometrics was used to predict the protein content for wheat (Williams, 2001). Since then chemometrics has found applications in many food and beverage related industries. In the wine industry it has been shown that, with the use of chemometrics, wines can be discriminated by cultivar, using Fourier transform infrared spectroscopy (FTIR) in conjunction with UV-visible spectroscopy and near infrared spectroscopy (NIR) in conjunction with visible spectroscopy (Edelmann *et al.*, 2001; Cozzolino *et al.*, 2003); as well as vintage using only FTIR (Palma and Barroso, 2002). Infrared spectroscopy has the advantage of being fast, non-destructive, and is particularly characterised by simplicity with regard to sample preparation.

For the first part of the study analysis were done using FTIR and GC. FTIR was chosen on the grounds that it can give a chemical fingerprint of the very complex matrix of a wine. As mentioned previously, analysis time and sample preparation for FTIR is also fast. GC was chosen on the basis that it has already been used to quantify aroma components in wine produced by different yeast species and strains (Romano *et al.*, 2003). Quantitative information from the GC can also be used for discrimination.

To discriminate between wines fermented with different yeast strains, a route of DNA extraction could be followed. Even though it is possible to isolate DNA from wine, it is quite difficult as wine is usually clarified and filtered (Ribéreau-Gayon *et al.*, 2000) largely reducing available DNA. There are further drawbacks with DNA extraction, as the extraction of DNA is a timely process (from overnight precipitation up to two weeks) (Savazzini and Martinelli, 2006). The extraction of DNA from wine is poor and amplification of DNA is difficult due to interference from tannins, polysaccharides and polyphenols present in the wine (Siret *et al.*, 2000; Savazzini and Martinelli, 2006).

FTIR has been used for the identification and discrimination of bacteria as far back as the 1950's and 1960's (Naumann *et al.*, 1991). With the advancement of infrared instrumentation, more powerful computers and advanced algorithms for multivariate data analysis and pattern recognition, FTIR as a tool has become widely accepted and used in the identification of microbes (Mariey *et al.*, 2001). FTIR can be seen as a rapid, whole organism fingerprint approach (Naumann *et al.*, 1991; Zhao *et al.*, 2006) that can be used in conjunction with chemometrics for identification purposes (Maquelin *et al.*, 2002). For reliable discrimination it is very important that FTIR measurements are reproducible and there are several factors that can influence this, including cell cycle, growth stage of the cells, growth conditions, sampling and sample preparation (Maquelin *et al.*, 2002).

Due the successful application of FTIR in the identification of microorganisms, it was of interest to see if FTIR could be used for discrimination of wine yeast strains used in the second part of the study. The study was conducted with the use of FTIR in transmission and ATR modes. In this study FTIR was used for the first time to discriminate between yeast strains used specifically for wine making.

In this exploratory study, the effectiveness of FTIR and GC in conjunction with chemometrics was assessed for its ability to discriminate wines that were fermented with different strains of *Saccharomyces cerevisiae*. The use of FTIR in transmission mode and in ATR mode was investigated to discriminate between wine yeast strains. In order to determine the influence of sample presentation, the yeast strains were prepared and presented to the instruments in different ways i.e., in liquid medium (yeast-peptone dextrose and water) and in dried form (granular, powder and pellet).

1.2 AIMS

The specific aims and approaches of this study were:

(i) to use PLS-discriminaton as a chemometric method;

(ii) to evaluate ability of mid infrared spectroscopy (MIR) and GC as instrumental techniques to discriminate between wines fermented with five (VIN13, WE372, VIN13-EXS, VIN13-PPK and ML01) different *Saccharomyces cerevisiae* yeast strains;

(iii) to compare the resulting discrimination using GC and MIR data respectively;

(iv) to evaluate the effectiveness of FTIR in transmission mode to discriminate between two *Saccharomyces cerevisiae* strains (VIN13 and WE372) suspended in YPD and water;

(v) to evaluate the effectiveness of FTIR in attenuated total reflectance (ATR) mode to discriminate between five *Saccharomyces cerevisiae* active dried wine yeast strains (ADWY) (Maurivin B, AWRI R2, NT7, and VIN13), presented to the ATR in granulated, powder and pellet form; and

(vi) to evaluate the effectiveness of FTIR in attenuated total reflectance (ATR) mode to discriminate between five *Saccharomyces cerevisiae* yeast strains, prepared from liquid cultures and presented to the ATR in powdered form.

1.3 LITERATURE SITED

Cozzolino, D., Smyth, H.E., Gishen, M., 2003. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. J Agric Food Chem 51 (26), 7703-7708

Edelmann, A., Diewok, J., Schuster, K.C., Lendl, B., 2001. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. J Agric Food Chem 49 (3), 1139-1145

Maquelin, K., Kirschner, C., Choo-Smith, L.P., van den Braak, N., Endtz, H.P., Naumann, D., Puppels, G.J., 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J Microbiol Methods 51 (3), 255-271

Mariey, L., Signolle, J.P., Amiel, C., Travert, J., 2001. Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. Vibr Spectrosc 26 (2), 151-159

Naumann, D., Helm, D., Labischinski, H., Giesbrecht, P., 1991. The characterization of microorganisms by Fourier-transform infrared spectroscopy. In: Nelson, W. H. (ed). Modern Techniques for Rapid Microbiological Analysis, John Wiley & Sons New York. pp. 263

Palma, M., Barroso, C.G., 2002. Application of FT-IR spectroscopy to the characterisation and classification of wines, brandies and other distilled drinks. Talanta 58 (2), 265-271

Reid, L.M., O'Donnell, C.P., Downey, G., 2006. Recent technological advances for the determination of food authenticity. Trends Food Sci Technol 17 (7), 344-353

Ribéreau-Gayon, J., Glories, Y., Maujean, A., Dubourdieu, D., 2000. The chemistry of wine stabilization and treatments. In: Ribéreau-Gayon, J. (ed). Handbook of Enology, 2. Wiley, Chichester, England. pp. 404

Romano, P., Fiore, C., Paraggio, M., Caruso, M., Capece, A., 2003. Function of yeast species and strains in wine flavour. Int J Food Microbiol 86 (1-2), 169-180

Savazzini, F., Martinelli, L., 2006. DNA analysis in wines: Development of methods for enhanced extraction and real-time polymerase chain reaction quantification. Anal Chim Ac 563 (1-2), 274-282

Siret, R., Boursiquot, J.M., Merle, M.H., Cabanis, J.C., This, P., 2000. Toward the authentication of varietal wines by the analysis of grape (*Vitis vinifera* L.) residual DNA in must and wine using microsatellite markers. J Agric Food Chem 48 (10), 5035-5040

Williams, P.C., 2001. Implementation of Near-Infrared Technology. In: Williams, P. C. and Norris, K. (ed). Near Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, St. Paul, Minnesota, USA. pp. 296

Zhao, H., Parry, R.L., Ellis, D.I., Griffith, G.W., Goodacre, R., 2006. The rapid differentiation of *Streptomyces* isolates using Fourier transform infrared spectroscopy. Vibr Spectrosc 40 (2), 213-218

Chapter 2

LITERATURE REVIEW

The use of chemometrics in oenology

2.1 INTRODUCTION

Naes *et al.* (2002) defines chemometrics as “the use of statistical and mathematical procedures to extract information from chemical (and physical) data.

Chemometrics is widely used in the food industry, but why? The answer to this question has to do with the advancement of technology. Due to increased computational capacity of modern digital computers and smaller and more reliable electronics, analytical instruments have become much more powerful. People struggle without the help of computers to interpret the mass of data generated by analytical instruments. Handling of these large amounts of data, which is mostly of a multivariate nature, is done by using chemometrics. An example would be a typical FTIR spectrum that can have over a thousand variables. There are several chemometric software packages which are also dependant on the advancement of digital technology to utilise the processing capacity to full measure.

The use of chemometrics in the food, feed and beverage sciences is closely related to the increased application of near infrared (NIR) spectroscopy in wheat industry, where it was already used in the early 1970’s (Williams, 2001). Since then chemometrics has found applications in many food related industries and with the use of many different types of analytical instruments. The field of chemometrics is constantly expanding with new techniques to improve survey data analysis, classification, prediction, discrimination or to improve pre-processing of data. Many of these chemometric advances can be found in the two pre-eminent journals in the field of chemometrics, namely *Journal of Chemometrics* (Wiley) and *Chemometrics and Intelligent Laboratory Systems* (Elsevier). Research results applying chemometrics to different food and beverage related areas can also be found in many subject-specific journals.

In the food and beverage industry, chemometrics in combination with various analytical techniques is used for authentication of products and prediction of quantitative parameters. Prediction of quantitative sample parameters is done through a multivariate calibration model using multivariate analytical output from an analytical instrument and is validated with known reference values. Calibrations must span the entire range of expected values for any sample that might be encountered in routine analysis (Wetzel, 1998). The predictive ability of a chemometric calibration would usually be used in quality control (QC) for a fast prediction of some or many quality related parameters that would normally involve long, destructive and complicated analysis.

Various analytical instruments are used in routine predictive analysis, but one of the most frequently used analytical techniques is NIR spectroscopy. NIR was used in the

wheat industry for prediction of protein content using a regression models, which has since become a kind of role model archetype for applied studies in multivariate calibration (Williams, 2001). This approach of prediction through a multivariate calibration is now part of many commercial NIR instruments supplied by Buchi, FOSS and other analytical instrument suppliers. NIR has also been used for prediction of water content in milk powder (Reh *et al.*, 2004), NaCl concentration in sausage (Ellekjær *et al.*, 1993), protein and other parameters in feed soybean (Edney *et al.*, 1994), to mention just a few NIR applications. Some other instruments used for quality control prediction are X-ray spectroscopy for measuring metals in tea (Manhas Verbi Pereira *et al.*, 2006). Mid-Infrared (MIR) spectroscopy to predict chemical parameters of European Emmental cheeses produced during summer (Karoui *et al.*, 2006) and FT-Raman spectroscopy for the simultaneous determination of fructose and glucose in honey (Batsoulis *et al.*, 2005).

Another area where chemometrics is frequently used is in authentication (Reid *et al.*, 2006). Authenticity is defined as “worthy of belief as conforming to fact or reality; trustworthy; genuine” (Longman, 1982). Generally, foodstuffs are either of animal or plant origin and for reliable authentication, parameters should be used that do not undergo significant alteration during processing (Luthy, 1999). In authentication studies the aim is to group products together in order to highlight the products that have been altered by addition of a cheaper, but similar substance; products that are mislabelled or to identify the origin of the products.

Authentication is a rapidly expanding area and is mainly driven by legislation that protects the status of products or where food safety is a concern. A summary of systems used in the European Union for protection of food product origin can be found in Reid *et al.* (2006).

A wide variety of analytical instrumentation is used in authentication of foodstuffs. For authentication purposes the speed of analysis is not always the most important factor. Some of the analytical instrumentation currently in use for authentication includes spectroscopy (UV, NIR, MIR, visible, Raman), isotopic analysis, chromatography, electronic nose, polymerase chain reaction, enzyme-linked immunosorbent assay and thermal analysis (Reid *et al.*, 2006), while a new modality would be the electronic tongue (Legin *et al.*, 2003). The theory and principles of the various analytical instruments falls outside this review, but can be found in *Principles of instrumental analysis* (Skoog *et al.*, 2007). Some recent reviews in the field of food quality control include Reid *et al.*, 2006); Martinez *et al.*, 2003); Tzouros and Arvanitoyannis, 2001); Wilson and Tapp, 1999) and Downey, 1998). A comprehensive

review regarding quality control methods for wine authenticity was done by Arvanitoyannis *et al.* (1999).

2.2 BASIC STATISTICS USED IN CHEMOMETRICS

The following section gives a brief description of some of the basic statistics concepts used in chemometric analysis of results as used in the review that follows. The description is by no means intended to be comprehensive. The formulas and descriptions dealt with in this section were assembled from chemometric textbooks by Esbensen (2002), Naes *et al.* (2002) and Spiegel (1972), respectively.

2.2.1 STANDARD DEVIATION, *s*

The standard deviation, *s*, is the root mean square of deviation from the mean of a set of n numbers, it is denoted by s and is defined by

$$s_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{(n-1)}} \qquad (2.2.1)$$

Where:

y_i is item *i* in the set

$\bar{Y}$ is the mean of the number set

2.2.2 ROOT MEAN SQUARE ERROR OF PREDICTION (*RMSEP*)

RMSEP is the prediction error estimate expressed in the original units of measure. RMSEP is defined as the square root of the mean of the squared differences between the predicted and measured reference values.

$$RMSEP = s_{Y.X} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad (2.2.2)$$

Where:

$\hat{y}_i$ is the predicted value for item *i* in the set

y_i is the measured reference value for item *i* in the set

n is the number of samples in the set

2.2.3 BIAS

Bias is the mean difference between the predicted and the measured reference values for all the samples in a validation set. Bias is a measure of the overall accuracy of a prediction model.

$$Bias = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \qquad (2.2.3)$$

Where:

$\hat{y}_i$ is the predicted value for item *i* in the set
y_i is the measured reference value for item *i* in the set
n is the number of samples in the set

2.2.4 STANDARD ERROR OF PREDICTION, *SEP*

SEP also known as *standard error of performance* is the standard deviation of residuals (difference between the predicted value and the reference value). It gives an indication of the variation of precision of the predicted values for several samples. It can also be described as the scatter around the regression line and is expressed as when corrected for bias

$$SEP = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i - Bias)^2}{(n-1)} \qquad (2.2.4)$$

Where:

$\hat{y}_i$ is the predicted value for item *i* in the set
y_i is the measured reference value for item *i* in the set
n is the number of samples in the set

If there is no bias, i.e. there are no differences between the mean values of the training and validation sets, the SEP is the same as the RMSEP.

2.2.5 COEFFICIENT OF DETERMINATION, r^2

The coefficient of determination is the ration of the explained variation to the total variation. If there is no explained variation the ratio is 0 and if all the variation is explained the ration is 1. In all other cases the ratio is between 0 and 1.

The correlation coefficient, *r*, is given by

$$r = \pm\sqrt{\frac{\text{explained variation}}{\text{total variation}}} = \pm\sqrt{\frac{\sum(Y_{pred} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}} \qquad (2.2.5)$$

Where:

Y_{pred} is the predicted value
$\bar{Y}$ is the mean of the number set
Y is the reference value

By substituting with equations for standard deviation (2.2.1), RMSEP (2.2.2) SEP, total variance (equation not shown), (2.2.3) can be written as

$$r = \sqrt{1 - \frac{s_{Y.X}^2}{s_Y^2}} \qquad (2.2.6)$$

It is clear that the coefficient of determination, r^2 then becomes

$$r^2 = 1 - \frac{s_{Y.X}^2}{s_Y^2} \qquad (2.2.7)$$

2.3 WIDELY USED CHEMOMETRIC TECHNIQUES

Prediction of quantitative parameters of samples is very important in QC. If a sample is analysed for only one property, say colour, at a single wavelength to predict ripeness, a univariate regression would be used to establish a calibration model for future prediction of ripeness. If, however, a single sample is analysed for many quantitative or qualitative parameters, for example the NIR spectrum of a sample to predict ripeness as well as many other properties, a multivariate calibration approach would be appropriate (Esbensen, 2002). Output from many modern analytical instruments, including those for on-line process analytical technology (PAT) purposes, demands use of multivariate regression models for future prediction (McLennan and Kowalski, 1995; Bakeev, 2005).

Unsupervised chemometric classification is made up of a group of techniques used to identify any internal data structure in a set; this general data analysis operation can be termed pattern cognition. These methods are used where there is no prior knowledge, or only very little knowledge available, pertaining to the data at hand. It is also used when a lot of information is available for a given data set, but to investigate any groupings and/or trends not hypothesised before. Generic cluster analysis is often done by principal component analysis (PCA, see below for explanation), or by

hierarchical techniques (also called cluster analysis methods) where a hierarchical pattern of distances between samples and agglomerated groups of samples are investigated to delineate patterns and clusters in the data set. Hierarchical techniques lead to dendograms which are a visual representation of the clustering process (Esbensen, 2002; Naes *et al.*, 2002).

Supervised classification (sometimes known loosely as discriminant analysis) performs a higher-level data analysis, pattern recognition, by which new samples are analysed regarding their similarity (or dissimilarity) with regard to a set of known classes (groups, clusters). Supervised techniques establish rules for when and how future unknown samples will be classified into such pre-determined classes.

Figure 2.1 outlines the chemometric methods discussed in this section.

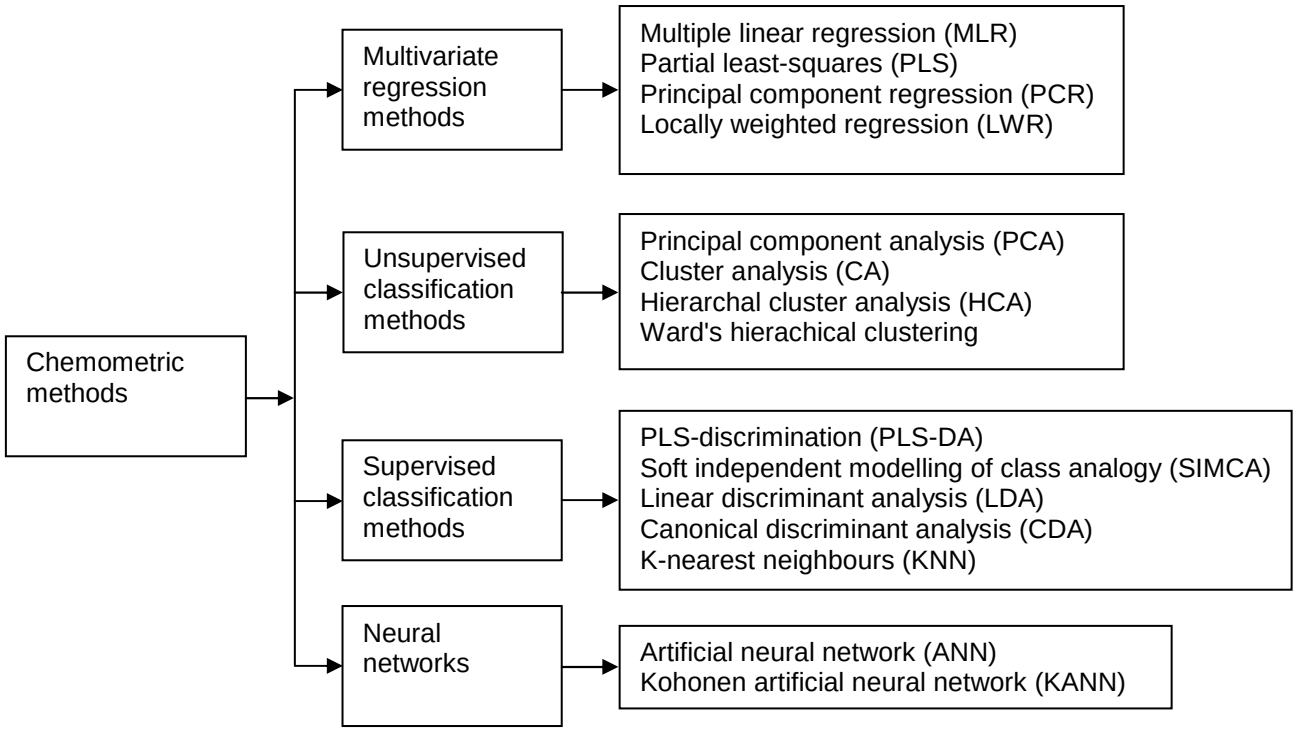


Figure 2.1: Outline of chemometrics methods discussed

2.3.1 MULTIVARIATE REGRESSION METHODS

2.3.1.1 Multiple linear regression (MLR)

MLR is an extension of a univariate regression with the difference being that in MLR one y-variable is regressed against several x-variables by least squares fitting (of the y-variable deviations). The critical drawback of this method is that all x-variables must be linearly independent, i.e. no significant X-variable collinearity is allowed. Outliers (those observations or variables which are abnormal compared to the major part of the data)

can also pose a serious threat to the accuracy of MLR (Esbensen, 2002; Naes *et al.*, 2002).

2.3.1.2 Partial least-squares (PLS) regression

PLS relates a single (also called PLS1), or many (also called PLS2), y-variables to a set of x-variables. In PLS regression each of the components or latent variables is calculated by maximising the covariance between the y-variable(s) and linear combinations of all x-variables, called the scores. In contrast to MLR, when employing a PLS method for regression, x-variable sets can show a high level of correlation or collinearity without the regression being affected, as is the case in many spectroscopic techniques. A set of PLS components is found, the first of these delineating that variation in the X-data which is most relevant to predicting the y-variable(s). The coordinates of the objects projected onto the new space are called scores. The loading weights of the X-variables signify how much each x-variable has in common with the y-variable for each component. As with PCA (see below), the scores and loading weights are usually presented graphically, presenting an optimised base for their interpretation (Esbensen, 2002).

2.3.1.3 Principal component regression (PCR)

PCR is a two step process, the first step which consists of a PCA on the x-variable set, to reduce dimensionality. In the second step, a standard MLR is performed using these principal component scores as the x-variable set (Esbensen, 2002).

2.3.1.4 Locally weighted regression (LWR)

LWR is used when dealing with non-linearities in data sets. LWR is based on PCR and assumes that there are local linearities in the data that can be utilised. For each new predicted sample, the x-variable set is projected down on the first couple of principal components (PC's). The calibration samples which are closest to the predicted sample are identified in this reduced dimensional space. Using only a few PC's and the local samples, a standard least squares solution is found. In this way a new calibration is performed on a local subset of calibration samples for each new prediction sample. As long as the number of samples in each local calibration and the number of PC's is small, there should not be any computational problems with this method (Naes *et al.*, 2002).

2.3.2 UNSUPERVISED CLASSIFICATION TECHNIQUES

2.3.2.1 Principal component analysis (PCA)

PCA optimally describes a data set in an original n-dimensional space by deriving a new set of underlying compound variables that are orthogonal to each other, while

minimising the loss of important data. The new variables can be thought of as linear combinations of all original X-variables. The first of these PC's is covering as much of the primary variation in the data as possible, with the second carrying the next highest fraction variance in a plane orthogonal to the first. The coordinates of the objects in the new space are termed object scores. Loadings are the coefficients by which the original variables must be multiplied to obtain the PC's. The numerical value of the loading is an indication of how much the variable has in common with a PC. The scores and loadings are usually graphically represented (Esbensen, 2002).

PCA is normally used to identify hidden patterns in a data set without knowing anything about the data beforehand. PCA is described as the "workhorse" of multivariate data analysis as almost all analysis is preceded, or should be, by a PCA to reveal possible data structure (Massart *et al.*, 1988; Esbensen, 2002).

2.3.2.2 Cluster analysis (CA)

Clustering in CA involves the measurement of either the distance or the similarity between objects (or variables). The distance measures selected are most often the Euclidean distance or Mahalanobis distance. The objects are then clustered in terms of their distance or similarity hierarchy (Naes *et al.*, 2002).

2.3.2.3 Hierarchal cluster analysis (HCA)

HCA groups objects in clusters on the basis of inter-object distances in high dimensional space. The results are shown in a dendrogram, which may be used to detect groups of similar individuals (Esbensen, 2002).

2.3.2.4 Ward's hierarchical clustering

Ward's method produces spherical clusters of roughly the same size. Using a pre-selected measure of similarity or distance, objects are clustered together. Starting with *n* groups each containing one object, this method is a so-called bottom-up approach. Two objects are combined to form a single cluster. A new object is then either added to the cluster or combined with another object to form a new cluster. This is continued until all objects belong to a cluster. Once a cluster is formed it cannot be split, it can only join with another cluster. Ward's method will join two groups when it will minimise the Error Sum of Squares. Due to the agglomerative nature of Ward's method, the cluster centres change each time a new object is added. This might mean that by the end of the process some objects are no longer in the correct cluster (Ward, 1963).

2.3.3 SUPERVISED CLASSIFICATION METHODS (DISCRIMINANT ANALYSIS)

2.3.3.1 PLS discrimination (PLS-DISCRIM, PLS-DA)

PLS-DA uses PLS regression to model the differences and thereby discriminate between classes (2 or more). This is done by assigning a dummy variable for each class. For a specific class a sample will be assigned +1 when it belongs to that class and -1 if it does not belong to that class. This system of +1 and -1 is used if there are only two classes and a PLS1 regression model is used. If there are more than two classes, a PLS2 regression will have to be used where each object has several dummy variables assigned to it, one for each class category. For example, if an object belongs to class 2 in a four-class problem it will have a variable set designation as follows: [-1;+1;-1;-1] (Esbensen, 2002).

2.3.3.2 Soft independent modeling of class analogy (SIMCA)

Soft independent modelling of class analogy (SIMCA) is a classification method based on individual PCA modelling of each class which can discriminated in the data. A PCA model is built on the training data for each known class of objects. Each PCA model will have its own optimum number of PC's as each class's data structure might be different from another. New samples are classified according to the class to which PCA model it fit best fits by calculating its distance to each PCA model in turn – then selecting the smallest. A new sample may also be classified as "not belonging to any" of the set of known classes; this option allows for detection of new types of samples, or of new aggregate patterns which is one of the most valuable assets of data analysis (Esbensen, 2002; Naes *et al.*, 2002).

2.3.3.3 Linear discriminant analysis (LDA), Canonical discriminant analysis (CDA)

LDA is very similar to CDA (also known as Fisher's Linear Discriminant analysis). LDA creates scatter plots from information found along the direction in multivariate space that separates groups as much as possible. Allocation rules can then be defined from the difference in groups. LDA first seeks a direction that maximises the difference between the groups' means as compared with the within-group variance. When there are only two groups this direction finding is the same for CDA and LDA. The line that defines the direction of maximum difference is called the canonical variate or linear discriminant function (LDF). CDA is used when there are more than two classes. The second LDF will describe the direction where the next best discrimination is and so on. For more than two classes the maximum number of LDF's is one less than the number of classes. The major drawback of these methods are that it assumes that covariance of the different classes are identical (Esbensen, 2002; Naes *et al.*, 2002).

2.3.3.4 K-nearest neighbours (KNN)

KNN classifies a new object by calculating its distance from each of the other objects in a training set. The K nearest neighbours (typical values for K are 3 or 5, chosen for performance optimisation) are found and the unknown is classified as belonging to the group that has the most members amongst these neighbours. This approach has the advantage of making no assumptions about the shapes of the groups at all. For more than two groups a tie-breaking situation might occur. An often used tie-braking rule is simply to use the nearest neighbour as indicator (Naes *et al.*, 2002).

2.3.4 NEURAL NETWORKS

2.3.4.1 Artificial neural networks (ANN))

An ANN consists of a nodes-net of information processing elements called neurones, which are connected together. They acquired ‘‘knowledge’’ by the calibration of the net, tested by the prediction of unknown input vectors which are not included in the calibration set. Generally, an ANN is organised into a hierarchy of layers: The first layer is the input layer with a node for each input variable, the output layer consist of a node for each variable to be determined – also encompassing a series of one or more hidden layers, between the input and the output layers, consisting of a given number of nodes. Each of the input nodes is connected to each of the hidden nodes and each of the hidden nodes is connected to each output node. Therefore, the signals are propagated from the input layer through the hidden layer(s) to the output layer. The contributions from all nodes are multiplied by constants (called weights) and added before the output of a node is determined by a nonlinear transfer function. Among the most popular nonlinear transfer function is the sigmoid function. The adequate functioning of a neural network strongly depends on the manner the signals are propagated through the net. The weights play a critical role in this propagation and a proper setting of these weights is essential. Usually, this setting is not known *a priori* and the weights are initially given randomly. The process of adapting the weights to an optimum set of values is called training, learning or calibration of the net. A representative training set is iteratively presented to the input of the neural network and the difference between the desired solution (target) and the net calculated one (output) is used to adapt the weights step-by-step, according to the learning algorithm. This difference, or error, is back-propagated from output to input of the network for a new iteration to correct the weights until the network error converges to an estimated level initially assigned (Naes *et al.*, 2002; Penza and Cassano, 2004a).

2.3.4.2 Kohonen artificial neural network (KANN)

The Kohonen artificial neural network (K-ANN), also known as the self organising map (SOM), is based on a non-interconnected, single layer of neurons, usually arranged in a two-dimensional hexagonal or rectangular grid. Responses are usually at the top of this grid. Underneath the top layer is a column of cells, each cell representing a descriptor. Each of the cells have a weight vector, the number of elements in this vector is equal to the number of variables in the input object (this can be a spectrum or chromatogram). The term "self-organising" refers to the fact that the map is trained without supervision. During the learning of the network, each sample from a predetermined training set is presented to the network in a random order. For each sample, the distance between the sample and every column of weights is calculated. The column with the minimum distance is considered the winning neuron. The weights of this neuron are modified so that at the following cycle the distance of the same sample from the winning neuron shall be smaller. A similar correction is applied to the neurons in the neighbourhood of the winner. This correction decreases with the distance. Usually the distance at which the correction takes place decreases during the learning phase. At the beginning the entire network is affected by every correction while in the last cycles only the winner neuron is corrected. Similarly, at the beginning, the learning rate and the amount of correction introduced is larger than in the latter cycles. The final result is a map, the first layer, where the most similar samples are in the same cell or next to one another. The weights give an insight into the reason for the clustering of the objects. Due of this, analysis of the first layer provides information on the similarity of the samples while the analysis of the weights provides information on the reason for their similarity (Kohonen, 1989; Marengo *et al.*, 2002).

2.3.5 VALIDATION

When using chemometric methods to predict quantitative parameters or when creating models for future discrimination of unknown samples, it is crucial to use proper model validation. Validation offers a prediction error estimate based on the calibration of a multivariate model. Proper validation of a multivariate model can also prevent over-fitting or under-fitting of data. There are three basic types of validation: leverage corrected validation, cross validation and independent test set validation. A comprehensive explanation of the types of validation is presented by Esbensen (2002). A short summary follows.

2.3.5.1 Independent test set validation

Independent test set validation is the best possible validation method to use in creating multivariate models (Esbensen, 2002). For these methods two completely independent

sets of data is required with known reference values. The two sets must be independent but similar with regards to processing conditions and the way the samples were taken and analysed. Both sets must be as similar, to any future samples that will be taken, as possible. The one set of data will then be used to create the calibration and the other to validate the model.

2.3.5.2 Cross validation

In this method of validation only one set of data is available for both the calibration and validation of a model. If only a few objects are available to build a calibration model then the so called leave-one-out or full cross validation is used. One object is taken out of a data set and the rest of the objects are used for the calibration model. The object left out is then used to validate the model. This process is continued until all objects were used as validation objects. The average of all the validation errors is then used as a measure of model accuracy.

It is obvious that full cross validation will result in an over optimistic model and that it might have no relation to any future data sets. This type of validation can also be extended into using segments of the data as calibration and validation sets. The optimum in accuracy being a two segmented cross validation where a data set is split in half and one half is used as calibration set and the other as validation set and then turning the two sets around (Esbensen, 2002).

2.3.5.3 Leverage corrected validation

This is a very quick and easy method, but results in a highly over optimistic model. Leverage measures the effect an individual object has on the model. The further an object is from the model centre the higher its leverage on the model. Leverage correction increases the weight of samples lying far from the model. Leverage correction is used early on in the modelling process when dealing with identifying outliers in the calibration data set. This method of validation should never be used for finalised models (Esbensen, 2002).

2.4 AUTHENTICATION IN THE FOOD INDUSTRY

As chemometrics has a big part of its origin in the food industry (Williams, 2001), a vast amount of research has been done in this field. This has led to the application of the same thought processes in other fields. What has been done in the food industry has had a big impact on how chemometrics has been applied in the wine industry. It is therefore imperative to look at some combinations of chemometrics and instrumental

analysis in the food industry. In the following paragraphs some of the instrumental and chemometric approaches used in the food industry are described.

In a recent review Reid *et al.* (2006) describe recent technological advances for the determination of food authenticity. The review covers various analytical instruments and accompanying chemometric evaluations. In a review by Fugel *et al.* (2005), quality and authenticity control of fruit purees, fruit preparations and jams with various analytical instruments and multivariate methods are discussed.

The effectiveness of the analysis of stable isotope ratios ($^{13}C/^{12}C$ and $^{15}N/^{14}N$) in fractions of lamb meat, measured by isotope ratio mass spectrometry was evaluated by Piasentier *et al.* (2003) as a method of authenticating feeding and geographical origin using canonical discriminant analysis (CDA). They were able to correctly classify 79.2% of samples based on country of origin and 91.7% based on feeding regime using cross-validation for both predictions.

Downey and Beauchene (1997) used NIR spectroscopy and selected chemometric techniques (PLS, FDA, SIMCA) to detect whether meat that has been frozen was substituted for fresh meat. Using meat drip samples that went through freeze-thaw cycles, they found that in a NIR spectral range of 1100 to 2498 nm the best separation was obtained by FDA.

Using chemical profiling methods, Anderson and Smith (2005) were able to determination the geographical origin (Iran, Turkey and USA) of pistachios. As part of the chemical profiling they made use of inductively coupled plasma atomic emission spectrometry (ICP-AES) for elemental analysis (Ba, Be, Ca, Cu, Cr, K, Mg, Mn, Na, V, Fe, Co, Ni, Cu, Zn, Sr, Ti, Cd, and P) and to analyse for inorganic anions and organic acids (selenite, bromate, fumarate, malate, selenate, pyruvate, acetate, phosphate, and ascorbate) they used capillary electrophoresis (CE). Bulk carbon and nitrogen isotope ratios were elucidated using stable isotope MS. The discrimination involved was achieved using CDA and PCA with accuracies of 95% and higher.

HPLC polyphenolic profiles of apple pulp, peel or juice provide enough information to develop classification criteria for establishing the technological grouping of apple cultivars (bitter or non-bitter) by using supervised pattern recognition procedures (LDA, KNN, SIMCA, PLS and multilayer feed forward ANN). In all cases for peel, pulp and juice 100% recognition and prediction were achieved (Alonso-Salces *et al.*, 2004).

Bortoleto *et al.* (2005) describes an innovative technique based on X-ray scattering applied to classify complex organic matrices of different vegetable oils. They used PCA

to discriminate between extra virgin olive oil from other olive oils and also to indicate the adulteration of extra virgin olive oil with soybean oil. The main reason for discrimination is attributed to the total lack of water in extra virgin olive oil.

Detection of Roundup Ready™ Soybeans by NIR spectroscopy with reasonable accuracy was achieved by Roussel *et al.* (2001). Chemometric techniques included Partial Least Squares (PLS), Locally Weighted Regression (LWR), and Artificial Neural Networks (ANN). Locally Weighted Regression using a database of approximately 8000 samples, provided the most accurate classification model (93% accuracy), while ANN and PLS methods provided classification accuracies of 88% and 78%, respectively.

The application of FTIR to identify possible adulteration of olive oils was adopted by Tay *et al.* (2002). Single-bounce attenuated total reflectance (ATR) measurements were made on pure olive oil as well as olive oil samples adulterated with varying concentrations of sunflower oil. Discriminant analysis was used to classify oil samples and PLS was used for the determination of concentration levels of the adulterant. Full cross-validation for the PLS model resulted in a R^2 of 0.974.

Karoui *et al.* (2005) investigated the potential of mid-infrared and intrinsic fluorescence spectroscopy for determining the geographic origin of different French and Swiss hard cheeses. By applying FDA to the MIR data only 80% correct classification was achieved. Using fluorescence spectroscopy 100% correct classification was achieved.

Determination of the geographic origin (Japan or China) of Welsh onions (*Allium fistulosum L.*) was conducted by Ariyama *et al.* (2004). They used flame atomic absorption spectroscopy, inductively coupled plasma atomic emission spectrometry (ICP-AES) and inductively coupled plasma mass spectrometry (ICP-MS) for elemental analysis of 20 elements (Na, P, K, Ca, Mg, Mn, Fe, Cu, Zn, Sr, Ba, Co, Ni, Rb, Mo, Cd, Cs, La, Ce, and Tl) together with LDA and SIMCA for classification. LDA provided a correct classification of 93% and SIMCA a correct classification of 91%.

A recent paper by Cordella *et al.* (2005), describes the development of an effective anionic chromatographic method (HPAEC-PAD) for honey analysis and the detection of adulteration with various industrial bee-feeding sugar syrups. Discrimination between authentic and adulterated honeys was done by LDA (96.5% correct classification) and to quantify the adulteration levels PLS analysis (R^2 using three components was 0.962) was used.

The above citations are only a few examples where different analytical instrumental output was combined with chemometric methods to authenticate food related products for various attributes.

In the sections that follow, an overview of the various combinations of instrumentation and chemometric approaches referred to in specific research articles related to wine will be given. Groups of research articles will be combined under a heading (in italic) that describes the type of instrument or instruments that were used in the research articles.

2.5 PREDICTION OF CHEMICAL PARAMETERS OF WINE

In the prediction of chemical parameters important to the oenological process, NIR and FTIR spectroscopy can be used. Both these types of analytical methods have the advantage of being fast in the analysis of a sample, typically less than 2 minutes per sample. The techniques are non-destructive and require minimal sample preparation, usually degassing and filtration only. Multiple chemical parameters are predicted simultaneously from the spectrum generated of the sample. The potential drawback of both these methods are that the predicted values are only as good as the models used, leaving the possibility for inferior calibration with the user. Upon reflection, this is a necessary prerequisite for any scientific endeavour of course – calibration must be the sole responsibility of the analyst/data analyst.

There are today excellent commercially available NIR and FTIR instruments which are dedicated to wine analysis in wine quality control laboratories. The Thermo Electron Corporation (Madison, Wisconsin) markets the Nicolet™ Antaris™ FT-NIR analyser. The analyser is able to predict properties for a wine, including density, ethanol content and °Brix. FOSS (Foss Electric, Denmark; http://www.foss.dk) markets the Winescan FT120 FTIR instrument. The Winescan comes pre-loaded with global calibrations for red, white and rose wines. The Winescan also offers the possibility for user created calibrations.

Some of the instrumental techniques and applications are discussed below.

NIR

Garcia-Jares and Médina (1997) used NIR reflectance with 19 interference filters for the simultaneous determination of ethanol, glycerol, fructose, glucose and residual sugars in botrytised-grape sweet white wines. By using a PLS model, predicted results compared well with other chemometric techniques like multiple linear regression (MLR), step wise regression (SWR) and principal component regression (PCR).

NIR spectroscopy *was* used by Cozzolino *et al.* (2004) to predict concentrations of malvidin-3-glucoside, pigmented polymers and tannins in red wine. They used 32 commercial red wines, totalling 495 samples, spanning two vintages, two grape varieties (Cabernet Sauvignon and Shiraz), two types of fermenters, two yeast strains and three different fermentation temperatures. A monochromator instrument was used to scan samples in transmission mode (400 to 2500 nm). The calibration was built using the NIR data as X and HPLC reference data (Y) with a PLS regression model; cross validation was used. A R^2 of greater than 0.8 was achieved. This was considered to be a rapid alternative method for prediction of red wine phenolics.

A feasibility study by Urbano-Cuadrado *et al.* (2004), they used NIR reflectance spectroscopy in wineries for determination of the 15 oenological parameters. The calibration and validation sets were built using 180 samples from six Spanish wine regions, three wine types, seven grape varieties and a mix of young and aged wines. Calibrations for ethanol, volumic mass, total acidity, pH, glycerol, colour, tonality and total polyphenol index was established using PLS regression and cross validation. R^2 values higher than 0.80 was achieved. Good correlations was also found for lactic acid, but less than desirable correlations for volatile acidity, malic acid, tartaric acid, gluconic acid, reducing sugars and SO_2 all with R^2 in the range 0.43 to 0.71.

Arana *et al.* (2005) were able to predict solids content of two varieties of Spanish grapes (Chardonnay and Viura) with NIR reflectance spectroscopy (800 to 500 nm). They found reasonable correlation coefficients, but that each variety needed its own PLS calibration using full cross-validation. Coefficients of determination for Chardonnay and Viura were 0.75 and 0.70 respectively.

NIR and FTIR

Urbano-Cuadrado *et al.* (2005) used NIR (400 to 2500 nm) and FT-IR (800 to 3000 cm^{-1}) independently and in combination to evaluate the prediction capability for several oenological parameters including alcoholic degree, volumic mass, total acidity, glycerol, total polyphenol index, lactic acid and total SO_2. It was found that NIR in general yielded better results, but when NIR and FTIR were combined, concentrations for glycerol and total SO_2 were even better determined. Calibrations were built using PLS regression and cross validation.

FTIR

Schneider *et al.* (2004) used FTIR to determine the glycosidic precursors responsible for varietal aroma in non-aromatic grapes. The only rapid test for glycoconjugates is the red-free glycosyl glucose method in which glucose is measured after acid hydrolysis, but

can only quantify total glycoconjugates. Samples (n=39) were collected at different maturity stages to be representative of the glycoside variability from Northwest France. Calibration models for the most relevant aroma glycoconjugates (C13-norisoprenoidic and monoterpenic glycoconjugates) for Muscadet wines were established using PLS regression with predictive errors of 14% and 15%, respectively.

Coimbra *et al.* (2005) found that by pre-treatment of FTIR spectra (1200 to 800 cm^{-1}) of red and white wine extracts with orthogonal signal correction (OSC) it was possible to quantify mannose polysaccharide from mannoproteins using PLS1 regression for calibration.

Nieuwoudt *et al.* (2004) developed a general calibration model with FTIR for predicting glycerol in wine (reducing sugar content < 30 g/L, alcohol > 8% v/v) with a SEP of 0.40 g/l. They further also developed a calibration model for special late harvest and noble late harvest wines (reducing sugar content 31-147 g/L, alcohol > 11.6% v/v) with a SEP of 0.65 g/L.

Various calibrations were developed by Urtubia *et al.* (2004) to monitor the complete fermentation process for glucose, fructose, glycerol, ethanol, malic acid, tartaric acid, succinic acid, lactic acid, acetic acid and citric by FTIR. The calibration models were built using PLS regression on Cabernet Sauvignon fermentations. Average error of prediction was 4.8% with malic acid the worst at 8.7%. Due to the low number of samples it was found that the calibrations were less good once external validation was used on fermentations of other varieties (test set validation).

Cocciardi *et al.* (2005) showed that single bounce attenuated reflectance (SB-ATR) FTIR performs better than FT-NIR and is comparable to transmission FTIR. PLS calibration (72 samples) with independent test validation (77 samples) using SB-ATR-FTIR for 11 wine parameters showed good correlation coefficients except for citric acid, volatile acid and SO_2.

E-nose

Maciejewska *et al.* (2006) showed that it is possible to follow a wine fermentation by using an array of partial selective gas sensors. They extracted the first PC and correlated it with ethanol content and volatile acidity with a high correlation coefficient and to a lesser degree also found a correlation with ethyl acetate. This study found a strong relationship between the first PC of the sensor array and human sensory patterns for the progression of the fermentation.

2.6 DISCRIMINATION OF WINES BASED ON ORIGIN

Sensory and analytical, (GC, HPLC)

Sivertsen *et al.* (1999) set out to discriminate wine (n=22) from four wine areas in France by using chemical analysis and sensory data. Chemical analysis was conducted by HPLC, GC and official analytical methods and included major acids, alcohols, esters, pH, total phenols and colour. Sensory analysis was done with 17 attributes. PCA was done and followed by CDA using the score matrix from the PCA to classify the wines in groups according to the four regions. It was found that the best classification was achieved by using chemical analysis data (81.8% correct classification) and the use of sensory data resulted in a distinctly less good classification (63.6% correct classification). The worse performance of the sensory data was attributed to a lack in good descriptors and an untrained panel.

Kallithraka *et al.* (2001) managed to classify 33 red Greek wines in two regions of Greece, Northern Greece and Southern Greece. They also used both chemical and sensorial data, but included mineral analysis by ICP as well. Using only PCA they could not discriminate between origin when using all instrumental and sensorial data. Clustering using PCA was successful when they used sensorial and anthocyanin data alone. The use of anthocyanins proofed to be a crucial factor in the discrimination of red wines while phenols and minerals were not as useful.

AA, analysis, chromatography, phenolic

By using stepwise linear discriminant analysis (S-LDA) on 12 analytical parameters for wine, Perez-Magarino *et al.* (2002) was able to classify rose wines into one of three Spanish protected designation of origin (PDO). They found that after samples were analysed for elemental composition by atomic absorption spectrometry, phenolics, colour measurements and classical wine parameters (ethanol, acidity), that ethanol and calcium were the most important parameters for discrimination as ranked by their statistical *F* values.

Arozarena *et al.* (2000) determined 20 analytical parameters for 66 wines making use of standard methods including GC-FID for volatile components. With the utilisation of factor analysis they were able to classify the wines into the two Spanish production areas where the originated from. By employing stepwise discriminant analysis they were able to get 92% of a proper test set of wines correctly classified.

Nuclear magnetic resonance (NMR)

Brescia *et al.* (2002) studied 41 red wines from Southern Italy. By using PCA and HCA they showed the presence of the three regional clusters and correct classification with DA on the two datasets of chemical data (chromatographic, routine analysis, ICP-AES)

and ^{1}H-NMR data. This showed that ^{1}H-NMR can be used for authentication of Italian wines in a much faster fashion than traditional techniques such as ICP-AES, routine analysis and chromatographic techniques.

NIR

Arana *et al.* (2005), used NIR (800 to 500 nm) reflectance data in combination with CDA to classify grapes of the same variety into two Spanish growing regions.

Mass Spectrometry (MS)

Headspace Mass Spectrometry (HS-MS) was applied by Marti *et al.* (2004) to red wines from two geographically close Catalonian growing areas. By variable selection they were able to show discrimination of the two growing areas using PCA and SIMCA.

Inductively coupled plasma (ICP)

Classification of 53 wines from four of the most important Bohemian (Czech Republic) regions according to their origin was undertaken by Sperkova and Suchanek (2005), with the use of ICP-MS and ICP-OES. The sample set consisted of red and white wines and 14 different varietals. In total 27 elements were determined. Clustering was observed when a PCA was performed on the elemental data and nearly 100% correct classification was achieved in a DA model for all regions. It was found that using the elemental variables represented by Al, Ba, Ca, Co, K, Li, Mg, Mn, Mo, Rb, Sr and V and element ratios Sr/Ba, Sr/Ca and Sr/Mg achieved best discrimination. They also found that discrimination was independent of vintage.

Coetzee *et al.* (2005) showed that ICP-MS in combination with stepwise discriminant analysis or pairwise discriminant analysis can successfully discriminate between wines from 3 South African wine regions. It was shown that by using 20 (Li, B, Mg, Al, Si, Cl, Sc, Mn, Ni, Ga, Se, Rb, Sr, Nb, Cs, Ba, La, W, Tl, and U) of the 40 analysed elements which showed a difference in their means across the three regions, discrimination was achieved.

Elemental analysis by ICP-MS and ICP-OES was used by Thiel *et al.* (2004) to discriminate between wines from four wine regions in Germany. Discriminant analysis was used with a training set of 88 known samples and this led to a correct classification of 88.6%. This was achieved after selection of variables (by means of analysis of variance) that is not influenced by climate and winemaking practices, namely As, B, Be, Cs, Li, Mg, Pb, Si, Sn, Sr, Ti, W, and Y.

ICP-OES, ICP-MS, IRMS, NMR, FTIR

Gremaud *et al.* (2004) used a variety of analytical instruments (ICP-OES, ICP-MS, IRMS, NMR and FTIR) to authenticate geographical origin of Swiss wines. It was

showed by using a LDA model that the sampled wines could be categorised into the four wine regions within Switzerland. It was also found that elemental analysis was independent of vintage.

E-tongue

The use of an electronic tongue was investigated by Riul *et al.* (2004) for the classification of red wines. By using the output from impedance spectroscopy with the electronic tongue in a PCA, they could discriminate 6 clusters of red wines by different producers. By employing an ANN on the sensing data they were able to discriminate different clusters even if different storage conditions were used.

Capillary electrophoresis (CE)

Nunez *et al.* (2000) made use of CE to discriminate between Spanish wines from three wine regions (Ribeira Sacra and 2 non-Ribeira Sacra). Discrimination could be achieved by the analysis of 25 wines using 6 trace metals (Na, K, Ca, Mg, Mn and Li) determined by CE. Various chemometric discriminating methods (LDA, KNN and SIMCA) gave roughly the same results of above 90% recognition of Ribeira Sacra wines.

FTIR

Picque *et al.* (2005) used FTIR on dried Gamay wine extracts to determine their origin from three French wine regions by applying PLS discrimination. On average 85% of the test set wines were correctly classified.

HPLC

Rodriguez-Delgado *et al.* (2002) used HPLC to discriminate between 55 red wine samples from five production areas on the Canary Islands (Spain). The differentiation was achieved using LDA.

GC-MS

Marengo *et al.* (2002) found that by applying PCA, HCA and KANN (Kohonen artificial neural network) to volatile components (analysed with GC-MS) from 68 Italian wine samples they could discriminate wines from different vintages.

2.7 DISCRIMINATION OF WINE BASED ON THEIR VARIETAL

Standard oenological analysis

Arozarena *et al.* (2000) used standard oenological parameters and GC-FID analysis in conjunction with stepwise discriminant analysis (SDA) to discriminate seven Spanish varietals (Graciano, Tempranillo, Grenache, Carignan, Merlot, Ruby Cabernet and Cabernet Sauvignon). SDA performed much better than factor analysis which gave no discrimination except for Ruby Cabernet. For wines from the test set, 85% were correctly classified using SDA.

A combination of conventional oenological parameters, alcohols and esters was used by Aleixandre *et al.* (2002) to discriminate four varietal wines from Spain. In total 91 wines were analysed from two vintages and represented Cabernet Sauvignon, Tempranillo, Monastrell, and Bobal varietals. Discriminant analysis (DA) using the first two discriminant functions correctly classified 100% of wines from the 1994 vintage and correctly classified 95% of the 1995 vintage.

NMR

Kosir and Kidric (2002) used the signals of seven amino acids in Slovenian white wines obtained form one-dimensional ^{1}H NMR to discriminate between varietals. They only analysed 10 wines from the varietals Chardonnay, Welsch Riesling, Sauvignon and Riesling. Samples were clustered using Ward's hierarchical clustering method with Euclidean distances.

PTR-MS

Four wine varietals (two red and two white) were discriminated by Boscaini *et al.* (2004) using a method based on proton transfer reaction mass spectrometry (PTR-MS). This method analyses the volatile organic compounds present in the headspace of the wine. PCA was used on MS data and this resulted in the clustering of the four varietals.

GC-FID, GC-MS

Pet'ka *et al.* (2001) used the volatile composition, determined by GC-FID and GC-MS to classify Slovak white wines. The three varietals investigated were Welschriesling, Grüner Veltliner and Müller Thurgau. A final discrimination was achieved using 10 volatile components and LDA as a chemometric tool.

NIR

NIR reflectance (800 to 500 nm) was used by Arana *et al.* (2005) to discriminate between two white varietals from Spain based on the analysis of their juice. They found it possible to discriminate Viura from Chardonnay grapes using Canonical discriminant analysis with overall success of 97.2%.

Vis-NIR

The combination of visible and NIR spectroscopy was used by Cozzolino *et al.* (2003) to discriminate between wines made from two varietals (Chardonnay and Riesling). They used a total of 269 samples divided in to sets, one used as a training set and the other as a validation set (test set). Initially outliers were removed by studying the results from PCA on all samples. PLS-DA correctly classified 100% of Riesling and 96% of Chardonnay samples from the validation set.

FTIR-ATR

By using FTIR-ATR, Edelmann *et al.* (2001) was successful in discriminating between 38 Austrian red wines (Cabernet Sauvignon, Merlot, Pinot Noir, Blaufränkisch (Lemberger), St. Laurent, and Zweigelt). By making use of dried phenolic extracts from the wines combined with a SIMCA classification 97% of the wines in their test set was correctly classified.

FTIR, E-nose, UV

Three fast analytical instruments (FTIR, E-nose and UV spectrometry) were compared by Roussel *et al.* (2003a) to discriminate white grape musts from France. A total of 138 samples were analysed from three varietals (Sauvignon, Mauzac and Colombard) and another group of mixed varietal musts. By applying PLS-DA to the data it was shown that FTIR analysis resulted in the best discrimination with 9% incorrectly classified, this after a Genetic Algorithm pre-processing of the data.

Roussel *et al.* (2003b) used data-fusion from aroma sensors, FT-IR and UV spectrometry based on Bayesian inference to classify white grape varieties. The 107 samples represented Sauvignon, Mauzac, Colombard and a fourth class made up of 35 samples of various other white varieties from the south of France. This approach delivered significantly better classification results that the individual instruments. The error in classification was 4.7% compared to 9.6% for FTIR, the best of the individual instruments using PLS-DA with full internal cross-validation.

E-tongue

Riul *et al.* (2004) used a conducting polymer based e-tongue to discriminate between red wines. Impedance spectroscopy was used as detection method and using PCA distinct clustering of the wines was observed. Using an ANN 100% recognition of the wines was achieved.

E-nose

Marti *et al.* (2004) used HS-MS as an electronic nose (e-nose) to discriminate between Catalonian red varietals. Wines made from Cabernet Sauvignon, Tempranillo and Merlot was used in the study. By using PCA on selected variables the found acceptable discrimination for the 1999 and 2001 vintage and partial overlap between clusters in 2000. The pattern of classification was also different for each vintage.

Penza and Cassano (2004a) used a four sensor e-nose attempting to classify nine Italian wines by using ionic conductivity, pH and alcohol content. They found that PCA did not give sufficient discrimination and with the use of ANN the also did not get a high

rate of correct classification. They concluded that the technique might be refined by using more parameters and increase the number of sensors in the e-nose.

A surface acoustic wave (SAW) sensor array, an e-nose, has been developed by Santos *et al.* (2005) in order to discriminate six Spanish wines coming from four different grape varieties (Tempranillo and Grenache, Airén and Malvar). LDA separated the clusters of different varietals and a probabilistic neural network (PNN) predicted with an error rate of 14%.

2.8 DETECTION OF ADULTERATION IN WINES

SNIF-NMR, IRMS

Ogrinc *et al.* (2003) used SNIF-NMR (site-specific natural isotopic fractionation nuclear magnetic resonance) and IRMS (isotope-ratio mass spectrometry) to test the detection power of isotopic measurements for detection of adulteration of wine ethanol with beet and sugar ethanol. They also found that proving the authenticity of a certain wine is only possible when the results are compared with a reference (non-adulterated) sample from a data bank of the same vintage from the same region.

Kosir *et al.* (2001) explore the use of SNIF-NMR and IRMS methods in combination with PCA, KANN (Kohonen artificial neural network) and cluster analysis for detection of chaptalisation of Slovak wines. They also compare chemometric methods to find the optimal one for the discrimination between groups of wines of natural and enriched wines. They found that the best discrimination was achieved using KANN after combining both methods.

In 1990, site-specific natural isotopic fractionation nuclear magnetic resonance (SNIF-NMR) was recognised by the European Community as an official method of displaying the characterisation of wine [The Commission of the European Communities 33 (1990) L 272].

NIR

NIR (1110 to 2500nm) was used by Sáiz-Abajo (2005) to detect adulteration of wine vinegar by ethanol and molasses vinegar. They examined the influence of Orthogonal Signal Correction (OSC) on the classification ability of models generated. Using 96 vinegar samples and by employing OSC the classification improved significantly against no pre-processing.

Flow injection analysis (FIA)

A FIA system to evaluate adulteration of alcoholic beverages by water and ethanol was investigated by Da Costa *et al.* (2004) based on the measurement of the Schlieren effect

(gradients of refractive index). Using SIMCA models they were able to identify 100% of laboratory adulterated samples and 95% of actual adulterated samples. The advantage of this system is that it does not use and reagents.

E-nose

Penza and Cassano (2004b) used the responses from a multisensor array from wines tested by headspace sampling to recognise adulterated Italian wines. Samples were adulterated with methanol, ethanol and other wines of the same colour. A multisensor array has been used to generate the chemical pattern of the volatile compounds present in the wine samples. These responses were analysed with PCA and ANN. The cross-validated ANN correctly classified 93% of samples.

2.9 DISCRIMINATING WINES ACCORDING TO VINTAGE

GC, ICP, amino acids

The analysis of Italian wines for amino acid content, volatile content and metal ions was used by Seeber *et al.* (19910 to discriminate between different vintages. The used PCA as a feature selection tool and then used LDA for discrimination with an overall successful classification rate of 97.2%.

Chemical analysis

Giaccio and Del Signore (2004) used chemical analysis of 17 different chemical parameters to classify wines from the Abruzzo region in Italy according to their vintage. Wine samples (156 samples) were collected over three vintages from different origins within the region. By using LDA they were able to classify the wines according their vintage with 100% success.

2.10 CONCLUSION

There is an important general lesson to be found in the above results. A wide swath of instrumental, chemical and physical approaches have been evaluated, many with excellent results, though by no means all, the unavoidable publication bias towards positive results notwithstanding. Many combinations of classical statistical methods of data analysis, as well as data analytical chemometric approaches, would appear to work well in general. However, it is easy to draw too hasty, positive conclusions when employing the optimistic cross validation approach (segmented or full i.e. leave-one-object-out alternatives), which will necessarily lead to an overly optimistic, and hence unrealistic, prediction performance assessment (Esbensen, 2002), the only responsible approach is proper test set validation.

It is clear that there are as many methods for the prediction of parameters and authentication of wine as there are analytical instruments available. As consumer awareness regarding food safety increase and producer protection of their processes spread, more methods for authenticity will become available. The methods of high value will be those that are fast and can be automated with little or no human intervention. With more requirements placed on food quality speed of analysis and the ease of data processing will become more important. With the advancements in analytical instrumentation and chemometric software this will be possible.

2.11 ABBREVIATIONS USED

ANN	artificial neural networks
ATR	attenuated total reflectance
CDA	canonical discriminant analysis
CE	capillary electrophoresis
CA	cluster analysis
E-nose	electronic nose
E-tongue	electronic tongue
EC	European Council
FIA	flow injection analysis
FTIR	Fourier transform infrared
FT-IR	Fourier transform infrared
FT-Raman	Fourier transform Raman
GC-FID	gas chromatograph – flame ionising detector
GC-MS	gas chromatograph – mass spectrometry
GC	gas chromatography
GM	genetically modified
GMO	genetically modified organism
HS-MS	headspace mass spectroscopy
HCA	hierarchal cluster analysis
HPLC	high performance liquid chromatography
HPAEC-PAD	high-performance anion exchange chromatography with pulsed amperometric detection
ICP	inductively coupled plasma
ICP-AES	inductively coupled plasma atomic emission spectrometry
ICP-MS	inductively coupled plasma mass spectrometry
ICP-OES	inductively coupled plasma–optical emission spectroscopy
IRMS	isotopic ratio mass spectrometry
KNN	K-nearest neighbours
KANN	Kohonen artificial neural network
LDA	linear discriminant analysis
LDF	linear discriminant function
LWR	locally weighted regression

MIR	mid-infrared
MLR	multiple linear regression
NIR	near infrared
NMR	nuclear magnetic resonance
OSC	orthogonal signal correction
PLS	partial least-squares
PLS-DISCRIM	PLS discrimination
PLS-DA	PLS discrimination
PCA	principal component analysis
PCR	principal component regression
PC	principal components
PDO	protected designation of origin
PDO	protected designation of origin
PGI	protected geographical indication
PTR-MS	proton transfer reaction mass spectrometry
QC	quality control
SOM	self organising map
SNIF-NMR	site-specific natural isotopic fractionation nuclear magnetic resonance
SIMCA	soft independent modeling of class analogy
SWR	step wise regression
S-LDA	stepwise linear discriminant analysis
TSG	traditional specialty guaranteed
UV	ultra violet
Vis-NIR	visible – near infrared

2.12 LITERATURE CITED

Aleixandre, J.L., Lizama, V., Alvarez, I., Garcia, M.J., 2002. Varietal differentiation of red wines in the Valencian region (Spain). J Agric Food Chem 50 (4), 751-755

Alonso-Salces, R.M., Herrero, C., Barranco, A., Berrueta, L.A., Gallo, B., Vicente, F., 2004. Technological classification of Basque cider apple cultivars according to their polyphenolic profiles by pattern recognition analysis. J Agric Food Chem 52 (26), 8006-8016

Anderson, K.A., Smith, B.W., 2005. Use of chemical profiling to differentiate geographic growing origin of raw pistachios. J Agric Food Chem 53 (2), 410-418

Arana, I., Jaren, C., Arazuri, S., 2005. Maturity, variety and origin determination in white grapes (Vitis Vinifera L.) using near infrared reflectance technology. J Near Infrared Spectrosc 13 (6), 349-357

Ariyama, K., Horita, H., Yasui, A., 2004. Application of inorganic element ratios to chemometrics for determination of the geographic origin of Welsh onions. J Agric Food Chem 52 (19), 5803-5809

Arozarena, I., Casp, A., Marín, R., Navarro, M., 2000. Multivariate differentiation of Spanish red wines according to region and variety. J Sci Food Agric 80 (13), 1909-1917

Arvanitoyannis, I.S., Katsota, M.N., Psarra, E.P., Soufleros, E.H., Kallithraka, S., 1999. Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). Trends Food Sci Technol 10 (10), 321-336

Bakeev, K., 2005, Process Analitycal Technology Spectroscopic Tools & Implementation. Blackwell Publishing, Oxford

Batsoulis, A.N., Siatis, N.G., Kimbaris, A.C., Alissandrakis, E.K., Pappas, C.S., Tarantilis, P.A., Harizanis, P.C., Polissiou, M.G., 2005. FT-Raman spectroscopic simultaneous determination of fructose and glucose in honey. J Agric Food Chem 53 (2), 207-210

Bortoleto, G.G., Pataca, L.C.M., Bueno, M.I.M.S., 2005. A new application of X-ray scattering using principal component analysis - classification of vegetable oils. Anal Chim Ac 539 (1-2), 283-287

Boscaini, E., Mikoviny, T., Wisthaler, A., Hartungen, E.v., Mark, T.D., 2004. Characterization of wine with PTR-MS. Int J Mass Spectrom 239 (2-3), 215-219

Brescia, M.A., Caldarola, V., De Giglio, A., Benedetti, D., Fanizzi, F.P., Sacco, A., 2002. Characterization of the geographical origin of Italian red wines based on traditional and nuclear magnetic resonance spectrometric determinations. Anal Chim Ac 458 (1), 177-186

Cocciardi, R.A., Ismail, A.A., Sedman, J., 2005. Investigation of the potential utility of single-bounce attenuated total reflectance Fourier transform infrared spectroscopy in the analysis of distilled liquors and wines. J Agric Food Chem 53 (8), 2803-2809

Coetzee, P.P., Steffens, F.E., Eiselen, R.J., Augustyn, O.P., Balcaen, L., Vanhaecke, F., 2005. Multi-element analysis of South African wines by ICP-MS and their classification according to geographical origin. J Agric Food Chem 53 (13), 5060-5066

Coimbra, M.A., Barros, A.S., Coelho, E., Gonalves, F., Rocha, S.M., Delgadillo, I., 2005. Quantification of polymeric mannose in wine extracts by FT-IR spectroscopy and OSC-PLS1 regression. Carbohydr Polym 61 (4), 434-440

Cordella, C., Militao, J.S.L.T., Clement, M.-C., Drajnudel, P., Cabrol-Bass, D., 2005. Detection and quantification of honey adulteration via direct incorporation of sugar syrups or bee-feeding: preliminary study using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and chemometrics. Anal Chim Ac 531 (2), 239-248

Cozzolino, D., Kwiatkowski, M.J., Parker, M., Cynkar, W.U., Dambergs, R.G., Gishen, M., Herderich, M.J., 2004. Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. Anal Chim Ac 513 (1), 73-80

Cozzolino, D., Smyth, H.E., Gishen, M., 2003. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. J Agric Food Chem 51 (26), 7703-7708

Da Costa, R.S., Santos, S.R.B., Almeida, L.F., Nascimento, E.C.L., Pontes, M.J.C., Lima, R.A.C., Simoes, S.S., Araujo, M.C.U., 2004. A novel strategy to verification of adulteration in alcoholic beverages based on Schlieren effect measurements and chemometric techniques. Microchem J 78 (1), 27-33

Downey, G., 1998. Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. Trends Anal Chem 17 (7), 418-424

Downey, G., Beauchene, D., 1997. Authentication of fresh vs. frozen-then-thawed beef by near infrared reflectance spectroscopy of dried drip juice. Lebensm-Wiss Technol 30 (7), 721-726

Edelmann, A., Diewok, J., Schuster, K.C., Lendl, B., 2001. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. J Agric Food Chem 49 (3), 1139-1145

Edney, M.J., Morgan, J.E., Williams, P.C., Campbell, L.D., 1994. Analysis of feed barley by near infrared reflectance technology. J Near Infrared Spectrosc 2 33-41

Ellekjær, M.R., Hildrum, K.I., Næs, T., Isaksson, T., 1993. Determination of the sodium chloride content of sausages by near infrared spectroscopy. J Near Infrared Spectrosc 1 65-75

Esbensen, K.H., 2002 (5th ed). Multivariate Data Analysis - in Practice CAMO Process AS, Oslo

Fugel, R., Carle, R., Schieber, A., 2005. Quality and authenticity control of fruit purees, fruit preparations and jams--a review. Trends Food Sci Technol 16 (10), 433-441

Garcia-Jares, C.M., Médina, B., 1997. Application of multivariate calibration to the simultaneous routine determination of ethanol, glycerol, fructose, glucose and total residual sugars in botrytized-grape sweet wines by means of near-infrared reflectance spectroscopy. Fresenius Z Anal Chem 357 (1), 86-91

Giaccio, M., Del Signore, A., 2004. Multivariate classification of Montepulciano d'Abruzzo wine samples according to vintage year. J Sci Food Agric 84 (2), 164-172

Gremaud, G., Quaile, S., Piantini, U., Pfammatter, E., Corvi, C., 2004. Characterization of Swiss vineyards using isotopic data in combination with trace elements and classical parameters. Eur Food Res Technol 219 (1), 97

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 (4), 501-514

Karoui, R., Bosset, J.-O., Mazerolles, G., Kulmyrzaev, A., Dufour, E., 2005. Monitoring the geographic origin of both experimental French Jura hard cheeses and Swiss Gruyere and L'Etivaz PDO cheeses using mid-infrared and fluorescence spectroscopies: a preliminary investigation. Int Dairy J 15 (3), 275-286

Karoui, R., Mouazen, A., Dufour, É., Pillonel, L., Picque, D., De Baerdemaeker, J., Bosset, J.-O., 2006. Application of the MIR for the determination of some chemical parameters in European Emmental cheeses produced during summer. Eur Food Res Technol 222 (1), 165-170

Kohonen, T., 1989 (3rd ed). Self-organization and associative memory. Springer-Verlag, New York

Kosir, I.J., Kidric, J., 2002. Use of modern nuclear magnetic resonance spectroscopy in wine analysis: determination of minor compounds. Anal Chim Ac 458 (1), 77-84

Kosir, I.J., Kocjancic, M., Ogrinc, N., Kidric, J., 2001. Use of SNIF-NMR and IRMS in combination with chemometric methods for the determination of chaptalisation and geographical origin of wines (the example of Slovenian wines). Anal Chim Ac 429 (2), 195-206

Legin, A., Rudnitskaya, A., Lvova, L., Vlasov, Y., Di Natale, C., D'Amico, A., 2003. Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception. Anal Chim Ac 484 (1), 33-44

Longman, 1982 (1st). Longman New Universal Dictionary. Longman Group Limited, Harlow, Essex, England

Luthy, J., 1999. Detection strategies for food authenticity and genetically modified foods. Food Contr 10 (6), 359-361

Maciejewska, M., Szczurek, A., Kerenyi, Z., 2006. Utilisation of first principal component extracted from gas sensor measurements as a process control variable in wine fermentation. Sens Actuators B Chem 115 (1), 170-177

Manhas Verbi Pereira, F., Rodrigues Pereira-Filho, E., Maretti Silveira Bueno, M.I., 2006. Development of a methodology for calcium, iron, potassium, magnesium, manganese, and zinc quantification in teas using X-ray spectroscopy and multivariate calibration. J Agric Food Chem 54 (16), 5723-5730

Marengo, E., Aceto, M., Maurino, V., 2002. Classification of Nebbiolo-based wines from Piedmont (Italy) by means of solid-phase microextraction-gas chromatography-mass spectrometry of volatile compounds. J Chromatogr A 943 (1), 123-137

Marti, M.P., Busto, O., Guasch, J., 2004. Application of a headspace mass spectrometry system to the differentiation and classification of wines according to their origin, variety and ageing. J Chromatogr A 1057 (1-2), 211-217

Martinez, I., Aursand, M., Erikson, U., Singstad, T.E., Veliyulin, E., van der Zwaag, C., 2003. Destructive and non-destructive analytical techniques for authentication and composition analyses of foodstuffs. Trends Food Sci Technol 14 (12), 489-498

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., 1988 Chemometrics: A Textbook. Elsevier, Amsterdam

McLennan, F., Kowalski, B.R., 1995 Process analytical chemistry. Chapman & Hall, London

Naes, T., Isaksson, T., Fearn, T., Davies, T., 2002 A User-friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, UK

Nieuwoudt, H.H., Prior, B.A., Pretorius, I.S., Manley, M., Bauer, F.F., 2004. Principal component analysis applied to Fourier transform infrared spectroscopy for the design of calibration sets for glycerol prediction models in wine and for the detection and classification of outlier samples. J Agric Food Chem 52 (12), 3726-3735

Nunez, M., Pena, R.M., Herrero, C., Garcia-Martín, S., 2000. Analysis of some metals in wine by means of capillary electrophoresis. Application to the differentiation of Ribeira Sacra Spanish red wines. ANLSCY 28 97-101

Ogrinc, N., Košir, I., Spangenberg, J., Kidrič, J., 2003. The application of NMR and MS methods for detection of adulteration of wine, fruit juices, and olive oil. A review. Anal Bioanal Chem 376 (4), 424-430

Penza, M., Cassano, G., 2004a. Chemometric characterization of Italian wines by thin-film multisensors array and artificial neural networks. Food Chem 86 (2), 283-296

Penza, M., Cassano, G., 2004b. Recognition of adulteration of Italian wines by thin-film multisensor array and artificial neural networks. Anal Chim Ac 509 (2), 159-177

Perez-Magarino, S., Ortega-Heras, M., Gonzalez-San Jose, M.L., 2002. Multivariate classification of rose wines from different Spanish protected designations of origin. Anal Chim Ac 458 (1), 187-190

Pet'ka, J., Mocák, J., Farka, P., Balla, B., Ková, M., 2001. Classification of Slovak varietal white wines by volatile compounds. J Sci Food Agric 81 (15), 1533-1539

Piasentier, E., Valusso, R., Camin, F., Versini, G., 2003. Stable isotope ratio analysis for authentication of lamb meat. Meat Sci 64 (3), 239-247

Picque, D., Cattenoz, T., Corrieu, G., Berger, J.L., 2005. Discrimination of red wines according to their geographical origin and vintage year by the use of mid-infrared spectroscopy. Sci Aliments 25 207–220

Reh, C., Bhat, S.N., Berrut, S., 2004. Determination of water content in powdered milk. Food Chem 86 (3), 457-464

Reid, L.M., O'Donnell, C.P., Downey, G., 2006. Recent technological advances for the determination of food authenticity. Trends Food Sci Technol 17 (7), 344-353

Riul, J.A., de Sousa, H.C., Malmegrim, R.R., dos Santos, J.D.S., Carvalho, A.C.P.L.F., Fonseca, F.J., Oliveira, J.O.N., Mattoso, L.H.C., 2004. Wine classification by taste sensors made from ultra-thin films and using neural networks. Sens Actuators B Chem 98 (1), 77-82

Rodriguez-Delgado, M.-A., Gonzalez-Hernandez, G., Conde-Gonzalez, J.-E., Perez-Trujillo, J.-P., 2002. Principal component analysis of the polyphenol content in young red wines. Food Chem 78 (4), 523-532

Roussel, S., Bellon-Maurel, V., Roger, J.-M., Grenier, P., 2003a. Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. J Food Eng 60 (4), 407-419

Roussel, S., Bellon-Maurel, V., Roger, J.-M., Grenier, P., 2003b. Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. Chemom Intell Lab Syst 65 (2), 209-219

Roussel, S.A., Hardy, C.L., Hurburgh, C.R.J., Rippke, G.R., 2001. Detection of Roundup Ready™ Soybeans by Near-Infrared Spectroscopy. Appl Spectrosc 55 (10), 1425-1430

Sáiz-Abajo, M.-J., González-Sáiz,José-María, Pizarro,Consuelo, 2005. Orthogonal signal correction applied to the classification of wine and molasses vinegar samples by near-infrared spectroscopy. Feasibility study for the detection and quantification of adulterated vinegar samples. Anal Bioanal Chem 382 (2), 412-420

Santos, J.P., Fernandez, M.J., Fontecha, J.L., Lozano, J., Aleixandre, M., Garcia, M., Gutierrez, J., Horrillo, M.C., 2005. SAW sensor array for wine discrimination. Sens Actuators B Chem 107 (1), 291-295

Schneider, R., Charrier, F., Moutounet, M., Baumes, R., 2004. Rapid analysis of grape aroma glycoconjugates using Fourier-transform infrared spectrometry and chemometric techniques. Anal Chim Ac 513 (1), 91-96

Seeber, R., Sferlazzo, G., Leardi, R., Dalla Serra, A., Versinig, G., 1991. Multivariate data analysis in classification of musts and wines of the same variety according to vintage year. J Agric Food Chem 39 1764-1769

Sivertsen, H.K., Holen, B., Nicolaysen, F., Risvik, E., 1999. Classification of French red wines according to their geographical origin by the use of multivariate analyses. J Sci Food Agric 79 (1), 107-115

Skoog, D.A., Holler, F.J., Crouch, S.R., 2007 (6th). Principles of instrumental analysis. Thomson, Brooks/Cole, Belmont, CA, USA

Sperkova, J., Suchanek, M., 2005. Multivariate classification of wines from different Bohemian regions (Czech Republic). Food Chem 93 (4), 659-663

Spiegel, M.R., 1972 Theory and problems of statistics in SI units. McGraw-Hill Publishing, New York, USA

Tay, A., Singh, R.K., Krishnan, S.S., Gore, J.P., 2002. Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. Lebensm-Wiss Technol 35 (1), 99-103

Thiel, G., Geisler, G., Blechschmidt, I., Danzer, K., 2004. Determination of trace elements in wines and classification according to their provenance. Anal Bioanal Chem 378 (6), 1630-1636

Tzouros, N.E., Arvanitoyannis, I.S., 2001. Agricultural Produces: Synopsis of Employed Quality Control Methods for the Authentication of Foods and Application of Chemometrics for the Classification of Foods According to Their Variety or Geographical Origin Crit Rev Food Sci Nutr 41 (4), 287- 319

Urbano-Cuadrado, M., Luque de Castro, M.D., Perez-Juan, P.M., Garcia-Olmo, J., Gomez-Nieto, M.A., 2004. Near infrared reflectance spectroscopy and multivariate analysis in enology: Determination or screening of fifteen parameters in different types of wines. Anal Chim Ac 527 (1), 81-88

Urbano-Cuadrado, M., Luque de Castro, M.D., Perez Juan, P.M., Gomez-Nieto, M.A., 2005. Comparison and joint use of near infrared spectroscopy and Fourier transform mid infrared spectroscopy for the determination of wine parameters. Talanta 66 (1), 218-224

Urtubia, A., Ricardo Perez-Correa, J., Meurens, M., Agosin, E., 2004. Monitoring large scale wine fermentations with infrared spectroscopy. Talanta 64 (3), 778-784

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58 236-244

Wetzel, D.L.B., 1998. Analytical near infrared spectroscopy. In: Wetzel, D. L. B. and Charalambous, G. (ed). Developments in Food Science: Instrumental Methods in Food and Beverage Analysis, 39. Elsevier Science, Amsterdam. pp. 141-194

Williams, P.C., 2001. Implementation of Near-Infrared Technology. In: Williams, P. C. and Norris, K. (ed). Near Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, St. Paul, Minnesota, USA. pp. 296

Wilson, R.H., Tapp, H.S., 1999. Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. Trends Anal Chem 18 (2), 85-93

Chapter 3

RESEARCH RESULTS

Discrimination between wines fermented by different yeast strains: a feasibility study comparing Fourier transform infrared spectroscopy with gas chromatography

Discriminating wines fermented by different yeast strains: a feasibility study comparing Mid Infrared spectroscopy with Gas Chromatography

C.D. Osborne[1], K.H. Esbensen[1,2], H.H. Nieuwoudt[1], P van Rensburg[1]

1. IWBT (Institute for Wine Biotechnology), Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa

2. ACABS (Applied Chemometrics, Analytical Chemistry, Bioenergy, Applied Biotechnology & Sampling) research group, Aalborg University Esbjerg (AAUE), Denmark

Abstract

Wines from three cultivars (Clairette Blanche, Pinotage and Cabernet Sauvignon) that were fermented by five *Saccharomyces cerevisiae* yeast strains (VIN13, WE372, VIN13-EXS, VIN13-PPK and ML01) were analysed using gas chromatography (GC) and Fourier transform infrared spectroscopy (FTIR). The data obtained were analysed by PLS-Discrim, a chemometric method. Using the data from both the analytical methods, discrimination was observed between wines fermented with different yeast strains in each of the two vintages (2005 and 2006) for all the cultivars. When combining the data from the two vintages no discrimination could be observed between the fermented wines. The discrimination of the fermented wines was found to be similar when using data from GC and FTIR, respectively. Since analysis with FTIR is considerably faster than analysis by GC it is recommended that FTIR be used for future studies of a similar nature. The combination into one sample set of wines fermented with commercial wine yeast strains and wines fermented from closely related wine yeast strains (the parental strain and two genetically modified versions thereof (VIN13, VIN13-EXS and VIN13-PPK)), wines fermented with closely related stains did not show good discrimination between each other. Discrimination was found between wines fermented with genetically modified (GM) wine yeast strains and those fermented with non-GM wine yeast strains. This was done on a limited number of yeast strains and a larger study is needed to confirm these results as this is the first study of this nature and differences seen could be as result of the different phenotypes.

Keywords: Wine discrimination; FTIR; Mid-infrared spectroscopy; GC-FID; Chemometrics; PLS-discrim; *Saccharomyces cerevisiae* yeast; GMO

3.1 INTRODUCTION

Today, consumers constantly want to know more about the food they consume. It is the consumers right to know about the nutritional value, the origin, the ingredients, possible allergens and many more facts about a specific food product's safety. These issues are not only limited to consumers, but are shared by many producers. Producers not only want to deliver the highest quality foods but also want to protect their products against unscrupulous producers who would want to misrepresent the food they are selling to gain economic advantage. This has led to a plethora of laws protecting food (Reid *et al.*, 2006). Most countries around the world have legislation in place to regulate additives, production processes, labelling and other aspects relating to food safety. A recent addition to legislation concerns the use of genetically modified foodstuffs (European Commission, 2003a, European Commission, 2003b).

To ensure adherence to all legislation and consumer demands for food safety the field of food authentication is growing rapidly. New analytical methods are continuously developed to prove adherence in a faster and more reliable way. The most rapid techniques include those that utilise near-infrared (NIR) and Fourier transform infrared (FTIR) technology. These methods are usually combined with chemometric techniques (Arvanitoyannis *et al.*, 1999; Tzouros and Arvanitoyannis, 2001; Reid *et al.*, 2006). Infrared spectroscopy, in its broader sense, is widely used in the food (Scotter, 1997; Wilson *et al.*, 2001; Cerna *et al.*, 2003) and beverage (Engelhard *et al.*, 2004; Duarte *et al.*, 2004) industries for routine quality control and authentication. In the wine industry it has been shown that wines can be discriminated by cultivar, using FTIR in conjunction with UV-visible spectroscopy and NIR in conjunction with visible spectroscopy (Edelmann *et al.*, 2001; Cozzolino *et al.*, 2003); as well as vintage using only FTIR (Palma and Barroso, 2002). Infrared spectroscopy has the advantage that it is fast, non-destructive, and is particularly characterised by simplicity with regard to sample preparation. An apparent disadvantage would be that it is an indirect analytical method and predicts only through chemometric multivariate calibration. In order to create a suitable calibration model, a representative calibration data set is required that covers the full range of all properties and features met with when making analytical chemical prediction based on (FTIR, NIR) spectra. The predicted values will always only be as good as the particular set of reference samples used to build the model. All these, and several other essential aspects of multivariate calibration (including the critical issue of proper validation), are covered adequately in standard textbooks (Martens and Naes, 1998; Esbensen, 2002).

It is well known that yeast has an important influence on wine quality and composition. Approximately 95% of sugar in must is converted to ethanol and CO_2, 1% is converted to cellular material and the remaining 4% to other end products (Boulton *et al.*, 1995). The major volatile products of fermentation are the alcohols, higher alcohols, esters, organic acids and aldehydes (Lambrechts and Pretorius, 2000). Gas chromatography (GC) in combination with a Flame ionising detector (FID) and mass spectrometry (MS) is used to identify and quantify the volatile products while some of the non-volatile components can be determined with high pressure liquid chromatography (HPLC) or other chromatography techniques. Under the same conditions each yeast species' growth is characterised by a specific metabolic activity that will lead to the final concentration of flavour compounds in the wine. Within a species different stains of yeast may also be characterised by specific metabolic activity (Romano *et al.*, 2003a) and the ratios of 2,3-butanediol and acetoin isomers were used to discriminate wines according to the yeast used in inoculated fermentations (Romano *et al.*, 2003b). Different phenotypes within non-*Saccharomyces* yeast species were discriminated by using the secondary products formed by the yeast during fermentation (Romano *et al.*, 1997). GC and HPLC analysis has the drawback of sample preparation and analysis time being time consuming.

Since the introduction of commercial genetically modified (GM) food and feed, legislation has followed and aimed to control all aspects of research, risk assessments, production, labelling and transport of gene-modified organisms (GMO) (Anonymous, 2003). The new legislation was a combined result from end-consumer pressure groups as well as environmental and health groups with the aim to provide a fully transparent overview of modern biotechnology. These groups demand that consumers be able to exercise their right to choose food on an informed and reliable basis. In order to comply with this type of legislation it is necessary to develop fast and reliable techniques to identify genetically modified events in plant and other products.

Due to international economic competition, through globalisation and a rapidly growing consumer demand towards higher quality wine, the ancient art of winemaking has entered the era of modern biotechnology. *Saccharomyces cerevisiae* strains are developed with bio-preservation abilities and to improve fermentation and processing of wine and must (Pretorius and Bauer, 2002). ML01 (Springer Oenologie, www.lesaffreyeastcorp.com) received GRAS (Generally Accepted as Safe) status on June 30th 2003 by the United States Food and Drug Administration. This signalled the first commercial release of a GM wine yeast. In the United States only voluntary labelling of GM foodstuff is required, while the EU-market has adopted a stringent *zero acceptance* policy (Anonymous, 2003). South Africa currently follow a similar GMO policy as the United States where labelling of GMO products are voluntary.

At the end of fermentation, wine is usually clarified from any residual yeast and other suspended matter and may also undergo filtration before bottling (Ribéreau-Gayon *et al.*, 2000). This will lead to a significant reduction of residual DNA, whether it originates from micro-organisms or from the grapes. It has been shown, however, that it is still possible to extract *Vitis vinefera* DNA from must and wine (Siret *et al.*, 2000; Faria *et al.*, 2000; Siret *et al.*, 2002; Garcia-Beneytez *et al.*, 2002; Leopold *et al.*, 2003; Savazzini and Martinelli, 2006). It has also been shown that *S. cerevisiae* DNA can be extracted from both must, young and older wine (Leopold *et al.*, 2003; Savazzini and Martinelli, 2006). Recovered yeast DNA were amplified by PCR, or real-time PCR, and compared to a reference standard (Leopold *et al.*, 2003; Savazzini and Martinelli, 2006). Real-time PCR is the technique of choice in quantification and specific detection of transgenic DNA (Higuchi *et al.*, 1992; Higuchi *et al.*, 1993; Klein, 2002). It has been proposed that the *S. cerevisiae* endogenous *ScRPS3* gene be applied as a genotype referee and gene copy number standard used for quantification with real-time PCR (Savazzini and Martinelli, 2006). This technique unfortunately has several drawbacks; extraction of DNA is a timely process (from overnight precipitation up to two weeks) (Savazzini and Martinelli, 2006) and it suffers from poor extraction and amplification of DNA due to interference from tannins, polysaccharides and polyphenols present in the wine or must (Siret *et al.*, 2000; Savazzini and Martinelli, 2006)

The objective of this study is firstly to evaluate ability of FTIR and GC as instrumental techniques to discriminate between wines fermented with different *S. cerevisiae* yeast strains. A second objective is to compare the resulting discrimination using GC and FTIR data respectively. The discriminatory technique used in this study will be PLS-discrim. This study is of an exploratory nature.

3.2 MATERIALS AND METHODS

3.2.1 CULTIVARS

Clairette Blanche, Pinotage and Cabernet Sauvignon were used to represent cultivars of both dry white and dry red wine styles. The Clairette Blanche grapes of both 2005 and 2006 harvest were sourced from the KWV Grondves farm outside Stellenbosch, South Africa. Pinotage grapes of the 2005 harvest were sources from the Stellenbosch region and for the 2006 harvest from the Zevenwacht Estate outside Kuilsriver, Stellenbosch area. Cabernet Sauvignon grapes were sourced from Fort Simon (Bottlery district, Stellenbosch) in 2005 and from the Stellenbosch University experimental farm, Welgevallen, in Stellenbosch in 2006.

3.2.2 YEAST

Five *S. cerevisiae* yeast strains (Tabel 3.1) were used from the freeze culture collection of the Institute of Wine Biotechnology (Stellenbosch University). Two strains are commercially available non-genetically modified (non-GM) strains (VIN13 and WE372). The two genetically modified (GM) strains are recombinant strains of VIN13. VIN13-EXS (Strauss, 2003), secretes enzymes (endo-β-1-4 glucanase coded by *END1*, endo-β-1-4 xylanase coded by *XYNC*) to enhance colour extraction while VIN13-PPK (Strauss, 2003) secretes enzymes (pectin lyase coded by *PELE*, polygalacturonase coded by *PEH1*) to enhance clarification and reduce viscosity. The commercial GM yeast, ML01 (currently only available in the United States of America and Canada), was also included in the study. ML01 performs alcoholic fermentation and malolactic fermentation without the need of lactic acid bacteria. The modified ML01 strain contains the *Schizosaccharomyces pombe* malate transporter gene and the *Oenococcus oeni* malolactic enzyme gene under the *S. cerevisiae PGK1* promoter and terminator signals. Simultaneous expression of both genes triggers complete L-malate utilisation by the recombinant *S. cerevisiae* strain.

All yeast strains were streaked out on yeast-peptone-dextrose (YPD) plates from freeze cultures and incubated at 30℃ for 48 hours. A single colony was picked and inoculated in a 10 mL YPD liquid culture. Liquid cultures were grown overnight at 30℃ on a rotating wheel. Cultures were then inoculated into 1 L YPD flasks and grown at 30℃ for 48 hours. Yeast was centrifuged at 8665xg (Sorvall RC 5C, Germany; GSA rotor) for 3 minutes. The supernatant was discarded and the pellet washed twice with deionised water. The washed pellet was re-suspended in deionised water. In order to ensure that yeast was inoculated into the must at a final concentration of ~$3x10^6$ cells/mL, the yeast suspension was divided into 50 mL aliquots.

3.2.3 MICROVINIFICATION

Standard wine making techniques were employed for microvinification.

3.2.3.1 Clairette Blanche

Grapes were crushed and de-stemmed. Free run juice was separated from berries in a basket press. SO_2 (from potassium metabisulfite, Laffort, France; 30 ppm final concentration) was added to the juice before the juice was left overnight at 15℃ for solids to settle out. Clear juice was siphoned off and split into volumes of 4 L aliquots for fermentation in 4.5 L glass bottles. Di-amino phosphate (DAP) (Laffort, France, 70 ppm final concentration) and yeast to a final concentration of ~$3x10^6$ cells/mL were added to each fermenter. During the 2006 harvest a commercial liquid pectolytic enzyme (Pectazina H, Dal Cin Gildo spa, Italy; final concentration 0.01 g/L) was added to half of

the fermentations done with VIN13, WE372, VIN13-EXS and VIN13-PPK. This was done to mimic the enzymes secreted by VIN13-EXS and VIN13-PPK. Fermentation was completed at 15℃ and the fermentation was tracked by daily weighing of each fermenter. Fermentation was considered complete once there was no weight loss after three consecutive days and confirmed by determining residual sugar using mid-infrared spectroscopy (Winescan FT120, FOSS Electric, Denmark). Upon completion of fermentation, bentonite (Microcol, Laffort, France; final concentration 0.7 mg/L) was added for protein stabilisation. Wine was cold stabilised at 4℃ for two weeks. After stabilisation and clarification, wine was racked off the bentonite and yeast lees and bottled in 750 mL bottles with screw cap closures and stored at 4°C. Replicate fermentations were done for each yeast strain (Table 3.1).

3.2.3.2 Pinotage

Pinotage grapes were crushed and de-stemmed. Free run juice was separated from berries and split in equal volumes and combined with equal weight of crushed berries. SO_2 (from potassium metabisulfite, Laffort, France; 30 ppm final concentration in 2005 and 2006), di-amino phosphate (DAP) (70 ppm final concentration in 2005 and 2006) and yeast to a final concentration of ~$3x10^6$ cells/mL were added to each fermenter. During the 2006 harvest a commercial colour extraction enzyme (Rapidase Excolor, DSM, Netherlands; final concentration 2 g/hL) was added to half of fermentations made with VIN13 and WE372 to mimic enzymes secreted by two of the GM yeasts (VIN13-EXS, VIN13-PPK). Primary fermentation was done in 10 L buckets at 25℃ until the sugar concentration in the must was less than 3℗. Grapes were pressed in a basket press and pressed wine was transferred to 4.5 L glass bottles fitted with airlocks. Fermentation was deemed complete when there was no weight loss for three consecutive days. During the 2005 harvest malolactic fermentation occurred naturally in some of the fermentations. During the 2006 harvest the wine was inoculated with a commercial malolactic starter culture (Lalvin VP-41, Lallemand, France, final concentration 1 mg/L) at the pressing stage. After completion of alcoholic and malolactic fermentation, free SO_2 levels were checked (KI/KIO_3 titration, Metrohm, Switzerland) and adjusted to 40 ppm (free SO_2). The wine was raked off lees and put into cold stabilisation (4℃). After cold stabilisation wine s were racked of the fine lees and bottled in 750 mL bottles with screw cap closures. Replicate fermentations were done for each yeast strain (Table 3.1).

3.2.3.3 Cabernet Sauvignon

Cabernet Sauvignon wine was prepared as described for Pinotage. The number of replicate fermentations for each yeast strain is shown in Table 3.1.

3.2.4 EXPERIMENTAL PLAN

Fig. 3.1 sets out the experimental plan for the study conducted during the 2005 and 2006 vintages.

Table 3.1: Summary of yeast strains used for fermentation, cultivars inoculated and the number of independent fermentations done for each cultivar.

***S. cerevisiae* strain used for fermentation**	**GM status**	**Genotype/Description**	**Source/Reference**	**Year**	**Cultivar inoculated**	**No of fermentations**
VIN13	Non-GMO	Commercial strain	Anchor Yeast Bio-Technologies (SA)	2005	Clairette Blanche	16
					Pinotage	16
					Cabernet Sauvignon	15
				2006	Clairette Blanche	16
					Pinotage	31
					Cabernet Sauvignon	32
WE372	Non-GMO	Commercial strain	Anchor Yeast Bio-Technologies (SA)	2005	Cabernet Sauvignon	8
				2006	Clairette Blanche	16
					Pinotage	31
					Cabernet Sauvignon	32
VIN13-EXS	GMO	*ura3::ADH1$_P$-MFα1$_S$-END1-TRP5$_T$-ADH1$_P$-XYNC-ADH2$_T$*	Strauss, 2003	2005	Clairette Blanche	10
					Pinotage	10
					Cabernet Sauvignon	10
				2006	Clairette Blanche	16
					Pinotage	16
					Cabernet Sauvignon	16
VIN13-PPK	GMO	*ura3::ADH1$_P$-MFα1$_S$-PELE-TRP5$_T$-ADH1$_P$-MFα1$_S$-PEH1- TRP5$_T$*	Strauss, 2003	2005	Clairette Blanche	10
					Pinotage	9
					Cabernet Sauvignon	10
				2006	Clairette Blanche	16
					Pinotage	16
					Cabernet Sauvignon	16
ML01	GMO	Commercial GM strain	Springer Oenology	2006	Cabernet Sauvignon	32

3.2.4.1 2005 vintage

Four yeast strains were used, VIN13, VIN13-EXS, VIN13-PPK, WE372. Fermentations for Clairette Blanche and Pinotage were done with VIN13, VIN13-EXS and VIN13-PPK. For both Clairette Blanche and Pinotage sixteen independent fermentations were done with VIN13, ten independent fermentations with VIN13-EXS and ten and nine independent fermentations respectively with VIN13-EXS. For Cabernet Sauvignon fifteen independent fermentations were done with VIN13, ten independent fermentations with VIN13-EXS and VIN13-PPK whereas eight fermentations were done with WE372.

3.2.4.2 2006 vintage

For Clairette Blanche four yeast strains (VIN13, VIN13-EXS, VIN13-PPK, WE372) were used. For each yeast strain, sixteen independent fermentations were carried out.

Furthermore, for each yeast strain used, half of the fermentations were treated with a commercial pectolytic enzyme preparation (see sections 3.2.3.1 and 3.2.3.2).

For Pinotage fermentations four yeast strains (VIN13, VIN13-EXS, VIN13-PPK, WE372 were used and for fermentations of Cabernet Sauvignon five yeast strains (VIN13, VIN13-EXS, VIN13-PPK, WE372, ML01) were used. Two independent sets of fermentations were carried out. The second set of fermentations for each cultivar was started two days after the start of the first set of fermentations. This second set of fermentations was used as an independent test set for chemometric validation purposes. The second set of fermentations was carried out under the same conditions as the first set of fermentations, i.e. yeast suspensions, enzymes, nutrient additions and further treatments during the fermentations. In order to simulate the secretion of the enzymes from the genetically modified yeasts, commercial pectolytic enzyme preparations were added to half of each set of fermentations done with a specific yeast strain (except for VIN13-EXS and VIN13-PPK).

3.2.5 WINE SAMPLING

The sampling procedure was carried out to ensure that samples were representative according to the Theory of Sampling (TOS). Sampling liquids, in this case wine, are several orders of magnitudes less complicated than with slurries (must, pulp) (Petersen *et al.*, 2005; Petersen and Esbensen, 2005).

The first set of Clairette Blanche samples for 2005 (sample set 2005A) was taken three months after completion of alcoholic fermentation and the second set of samples (sample set 2005B) was taken eight months after completion of alcoholic fermentation. In 2006 the first set of samples (sample set 2006A) was taken two months after completion of alcoholic fermentation.

The first set of Pinotage samples in 2005 (sample set 2005A) was taken four months after completion of alcoholic fermentation and the second set of samples (sample set 2005B) was taken nine months after completion of alcoholic fermentation. In 2006 the first set of samples (sample set 2006A) was taken one week after completion of alcoholic fermentation, while the second set (sample set 2006B) was taken one month after completion of alcoholic fermentation and the third set (sample set 2006C) was taken two months after completion of alcoholic fermentation.

The first set of Cabernet Sauvignon samples (sample set 2005A) was taken three months after completion of alcoholic fermentation and the second set (sample set 2005B) was taken seven months after completion of alcoholic fermentation. In 2006 the

first set of samples (sample set 2006A) was taken two months after completion of alcoholic fermentation, while the second set (sample set 2006B) was taken two and a half months after completion of alcoholic fermentation and the third set (sample set 2006C) was taken four months after completion of alcoholic fermentation.

Samples were taken at different times during the ageing of the wines in order to study the effect that the ageing of a developing wine has on the ability to discriminate the wines fermented with different yeast strains. For red wines, the first set of samples was taken before the inoculation of malolactic starter cultures after pressing but before, in some instances, the completion of alcoholic fermentation. This resulted in varying levels of residual sugars. The last sample sets taken for the 2006 red wines were taken after the completion of malolactic fermentation (MLF).

After vigorous shaking, a 50 mL sample was extracted from the centre of each bottle and transferred to a 50 mL Corning tube. If samples were not analysed immediately they were stored at 4℃ for no mare than a month.

Samples (50 mL each) were taken from each of the fermentation replications for each of the fermentations done with the different yeast strains.

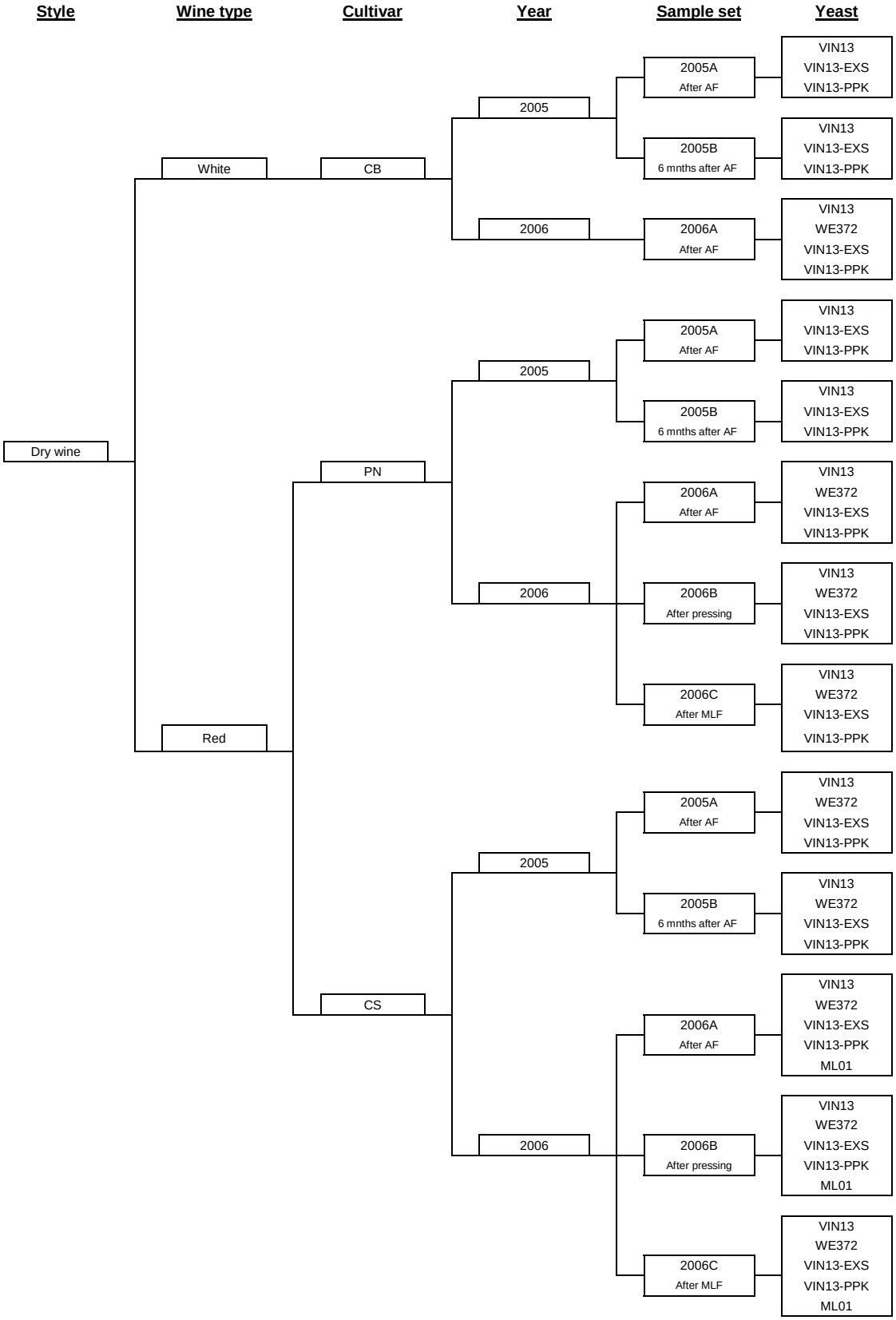


Figure 3.1: Experimental layout

CB: Clairette Blanche; PN: Pinotage; CS: Cabernet Sauvignon; AF: Alcoholic fermentation

3.2.6 INSTRUMENTAL

3.2.6.1 Fourier transform infrared (FTIR) analysis

Infrared analysis was carried out in the mid-infrared region (MIR, 5011 to 929 cm^{-1} at 4 cm^{-1} intervals) with a WineScan FT120 instrument (Foss Electric, Denmark). All samples were degassed to remove excess CO_2. Each sample was analysed in duplicate, each spectrum composed by the average of twenty scans after being Fourier transformed. Foss Zero Liquid S-6060 (WineScan FT120 Reference Manual) was analysed before wine samples to facilitate correction for the specific background spectrum (especially water) present in the wine matrix. The final spectra were generated based on the ratio of the individual sample spectrum to the zero solution spectra at each recorded wavelength. FTIR data are recorded as absorbencies as a function of wavenumber (cm^{-1}).

3.2.6.2 Gas chromatography (GC-FID)

GC-FID analysis was done on a HP 6890 Series (Agilent Technologies, USA) instrument. Samples were analysed for the 25 compounds listed in Table 3.2. Samples for the GC-FID were prepared by extracting 5 mL wine with 1 mL diethyl ether (Merck, South Africa) after adding 10 mg/L internal standard (4-methyl-2-pentanol, Fluka, Sigma-Aldrich, South Africa). Samples were sonicated for five minutes before centrifugation for three minutes at 4000 rpm (Heraeus MULTIFUGE 3S, Germany). Organic phase were dried on anhydrous sodium sulphate before transferred to a vial insert and capped. The GC-FID program used is shown in Table 3.3. Samples were analysed in triplicate and integrated on HP ChemStation (Rev A.07.01, Hewlett-Packard 1999) software.

Table 3.2: List of components analysed by GC-FID

2-Phenyl Ethanol	2-Phenylethyl Acetate	Acetic Acid	Butanol
Butyric Acid	Decanoic Acid	Diethyl Succinate	Ethyl Acetate
Ethyl Butyrate	Ethyl Caprate	Ethyl Caprylate	Ethyl Hexanoate
Ethyl Lactate	Hexanoic Acid	Hexanol	Hexyl Acetate
Isoamyl Acetate	Isoamyl alcohol	Isobutanol	Iso-Butyric Acid
Iso-Valeric Acid	Methanol	Octanoic Acid	Propanol
Propionic Acid	Valeric Acid		

Table 3.3: GC instrumental parameters

Oven Program	
Initial temperature	33℃
Initial time	17 min
Ramp	12℃/min to 240℃, hold for 5min
Front inlet	
Injection volume	3µl
Mode	Split
Split ratio	15:1
Split flow	49.5ml/min
Injection temperature	200℃
Initial pressure	84.5kPa
Flow mode	Constant flow
Column flow	3.3ml/min
Column	
Type	DB-FFAP, 60mx0.32mmx0.5umf.t
Detector	
Temperature	250℃
H_2 flow	30ml/min
Air flow	350ml/min
Make up flow	N_2 30ml/min

3.2.7 CHEMOMETRIC DATA ANALYSIS

Corrected (time and date stamps and other non-spectral data were removed from the raw data) FTIR spectra were imported in the Unscrambler software (version 9.2, Camo PROCESS AS, Oslo, Norway) for PLS-discrimination.

Pre-processing was carried out on all spectra. The second derivative was employed using the Savitzky-Golay algorithm (Savitzky and Golay, 1964) with five data points left – and right, using a second-order polynomial fit. The second order derivatives were used to compensate for baseline shifts, to remove possible minor scattering effects and for peak sharpening to bring out small peaks and changes in slope. Finally the data were mean centred and standardised by dividing each column in the data matrix with the standard deviation of each variable (wavenumber). This combined treatment is termed auto-scaling in chemometrics (Esbensen, 2002).

Although a spectral range of 930 to 4992 cm^{-1} was recorded, a reduced spectral range was used (965 to 1582 cm^{-1} and 1698 to 2971 cm^{-1}). The two regions, 1582 to 1698 cm^{-1} and 2971 to 3627 cm^{-1} contain water absorption bands (Patz *et al.*, 2004) while the region from 3627 to 4992 cm^{-1} is close to the near infrared region and lead to a noisy signal in the Winescan.

3.2.7.1 PLS2-discriminant analysis

PLS2-discriminant analysis was used for discriminating between fermentations with the different yeast strains. A dummy classification variable, Y, is created representing each of the five yeast strains used in the fermentations. A value of +1 was assigned to a sample when it was fermented with a specific yeast, and -1 if the specific yeast was not used. An example would be where a wine is fermented with VIN13, but not by WE372, VIN13-EXS, VIN13-PPK or ML01 then the Y-variable vector would be [+1,-1,-1,-1,-1]. PLS2-discrimination models are created for each sample set, set out in Fig. 1, with this classification variable set serving as five Y-variables throughout (or reduced when the discrimination issues are simpler, see below for each model). The X-variables for all these models were either the pre-treated FTIR spectra or GC data respectively. These PLS2-discrimination models are used to analyse the data structure of the experimental data. If more robust prediction models are needed in future for wines fermented with each individual strain, PLS1-discrimination models need to be created.

Contrary to conventional PLS-regression models in which the Y-variables are on a rational scale (so-called numeric variables, e.g. analyte concentration), for PLS-discrim models we are often only interested in the so-called “Predicted vs. Measured plot” because of the visual gap between predicted values centred on +1 and -1 respectively. The conventional prediction statistics (slope, r^2, RMSEP) have little meaning in this case (the prediction “model” is essentially a straight line connecting two points), these statistics are only of value when the reader has direct visual access to the Predicted vs. Measured plots. It is the gap centred on “0” which carries the essential discrimination information related to the classes modelled by the dummy Y-variables. This gap must be centred on “0”, lest the Predicted vs. Measured plot has been corrupted (often by ill-informed alternative scaling and/or specific pre-transformations) (Personal communication, Esbensen, K.H., 8 August 2007).

Interest is often only in the specific decomposition of the X-space alone, as guided by the Y-discrimination information assigned via the set of dummy variables (Esbensen, 2002). In this case, where there is no interest in prediction, no serious validation of the model is necessary and a simple leverage correction can be used. When interpreting X-space score plots, the score unit employed corresponds to a standard deviation of the variance along the specific component in question (Esbensen, 2002).

3.2.7.2 PLS1-discriminant analysis

PLS1-discriminant analysis was used to discriminate between non-genetically modified (non-GM) yeast fermentations and genetically modified (GM) yeast. A dummy variable was created representing either commercial non-genetically modified (non-GM)

yeast (signified by: -1) and genetically modified (GM) yeast (signified by: +1). These dummy variables were then modelled on the FTIR spectra (X-variables). A separate PLS-discrimination model was created for each sample set as set out in Fig. 3.1.

3.2.7.3 Chemometric output

Focus will only be on the specific discrimination features in the model documentations, especially the so-called “scores plot” and the “predicted vs. measured” validation plot.

The scores plot is a projection onto a particular sub-space, allowing optimal appreciation of the inter-sample relationships in two- or thee dimensions instead of the full variable FTIR-space (or GC-space). In this plot, assessment of the discrimination between the samples take place in precisely the two or three dimensions representing the largest variance differences between all samples, which is also maximally correlated to the [-1, +1] dichotomy in the Y-space, (Esbensen, 2002).

The “predicted vs. measured” validation plot summarises the prediction performances of the particular PLS-models. When a significant, centred, gap in the Y-direction has been obtained by a properly validated (prediction, the larger this gap, the more consistent and reliable the discriminations.

Loading plots were not evaluated due to the pre-processing of the FTIR spectra. The fact that PLS2-discrimination was used meant that a high loading at a specific wavenumber or GC component for one Y-dummy variable might have a low loading for another Y-dummy variable. In order to optimise variable selection for wine fermentations with a specific yeast, a PLS1-discrim analysis should be performed.

Some of the other features that can be visualised as well are the fractions X- and Y-variance modelled respectively. These graphical illustrations are the result from validation of the prediction power of the PLS-discrim models; validation is also used for determination of the optimal number of PLS-components in the model. Validation of a PLS-model forms a major issue in chemometrics, this merits serious attention and is discussed in full detail in Esbensen, 2002 The only two acceptable validation procedures for reliable assurance regarding prediction consistency and significance are:

1) 2-segment cross-validation (simulation of test set validation based on a large training set)
2) Test set validation (universally the optimal prediction validation)

Full description of PLS-regression can be found in the pre-eminent chemometric textbook of Martens and Naes (1998).

Each analysis of individual sample sets in the results section will indicate the type of validation used.

3.3 RESULTS AND DISCUSSION

3.3.1 DISCRIMINATION BETWEEN WINES FERMENTED WITH DIFFERENT YEAST STRAINS

Results from each cultivar will be presented to allow tracking of discrimination from a single sample set for a year through to looking at the cultivar over the two vintages. Each individual sample set of wines within a cultivar over the two vintages was evaluated to establish if any discrimination could be found between the wines that were fermented with the different yeast strains. Then the sample sets for each vintage was combined and evaluated to establish if the same discrimination can be observed that was seen in the separate sample sets within the singular vintage. This would indicate if changes in composition of the wine as it aged would change the discrimination of yeast strains. Lastly the two vintages will be combined to evaluate if the discrimination patterns, if any, follow through vintages. If the yeast strains can be discriminated for the sample set that span the two vintages, it could possibly be expanded to cover more vintages.

Due to the fact that fermentations done in the 2005 vintage was of an exploratory nature to establish if any discrimination was possible for the different yeast strains, no test set of fermentations was available here. This exploratory nature of the data analysis therefore has no need for a predictive function to test the strength of models created. In this particular case leverage correction was therefore used as validation method, but only the score plots are shown in the results below. For the 2006 vintage, two sets of independent fermentations were done with Pinotage and Cabernet Sauvignon, one set of which to be used as test set during validation. As FTIR and GC were used to obtain chemical fingerprints of the wine matrix no attention was given to the interpretation of the spectra or individual chromatograms.

3.3.1.1 Clairette Blanche

Fig. 3.2 represents a summary of the experimental layout for the wines made with the cultivar Clairette Blanche. It covers the two vintages, the independent sample sets drawn from the wines and the yeast used to ferment the musts with. In none of the sample sets an independent set of fermentations were produced that could serve as a test set for validation. Leverage correction was used in all the PLS2 discrimination

models, but below only the resulting full-model score plots are presented. A summary of results are presented in Table 3.4 at the end of the section.

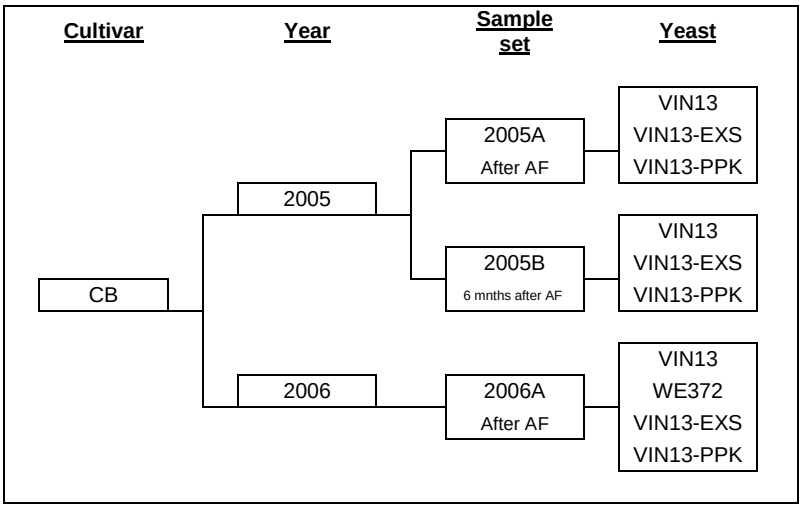


Figure 3.2: Experimental layout for wines made with cultivar Clairette Blanche.

3.3.1.1.1 Discrimination of wines in sample set 2005A

The sample set was analysed by FTIR. The FTIR data was used as X-variables. Fig. 3.3 shows the resulting score plot.

Three clusters of objects that represent wines fermented with the three yeasts used can be distinguished in tight clusters spread far apart.

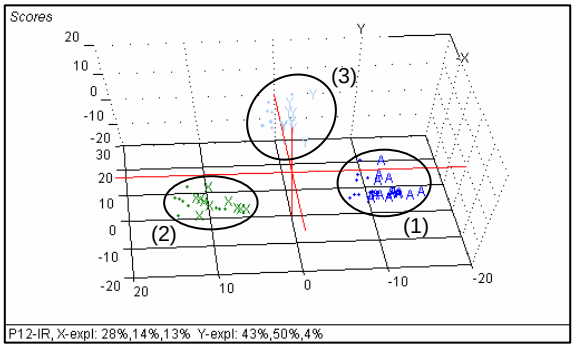


Figure 3.3: Score plot for Clairette Blanche sample set 2005A using only FTIR data.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

3.3.1.1.2 Discrimination of wines in sample set 2005B

The sample set was analysed using FTIR spectroscopy and gas chromatography.

Fig. 3.4a shows the score plot of the data set where FTIR data were used. Three tight clusters spread far apart can be seen representing wines fermented with the three different yeast strains.

Fig. 3.4b shows the score plot of the data set where only GC data was used. Three clusters can be seen.

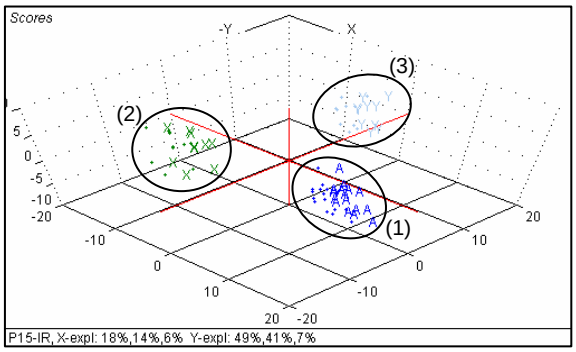


Figure 3.4a: Score plot for Clairette Blanche sample set 2005B using only FTIR data.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

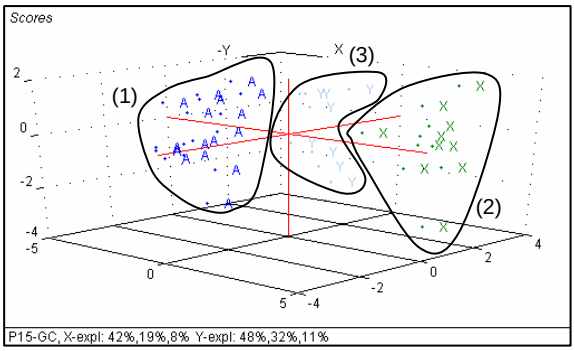


Figure 3.4b: Score plot for Clairette Blanche sample set 2005B using only GC data.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

3.3.1.1.3 Discrimination of wines in sample set 2006A

In 2006 fermentations were done with an extra yeast (WE372). The sample set was analysed using FTIR spectroscopy and gas chromatography.

The score plot using FTIR data is shown in Fig. 3.5a. Four tight clusters can be seen representing wines fermented with the four yeast strains used.

Using GC data, the score plot can be seen in Fig. 3.5b. Four tight clusters can be seen representing wines fermented with the four yeast strains used.

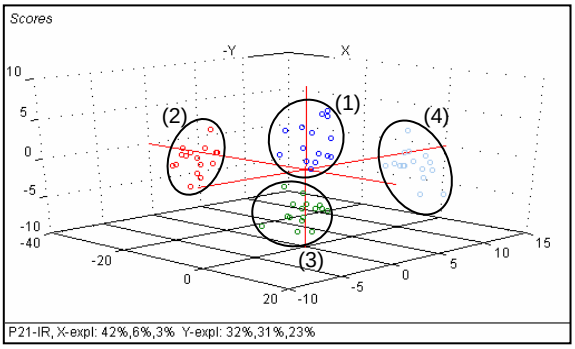


Figure 3.5a: Score plot for Clairette Blanche sample set 2006A using only FTIR data.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (3);
o - VIN13-PPK (4)

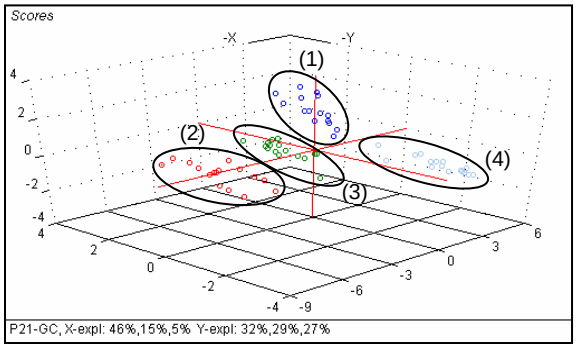


Figure 3.5b: Score plot for Clairette Blanche sample set 2006A using only GC data.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (3);
o - VIN13-PPK (4)

3.3.1.1.4 Discrimination of wines for combined 2005 data

The score plot in Fig. 3.6 show the combined data from sample sets 2005A and 2005B using FTIR data and therefore looking at the effect of ageing across the vintage. Three distinct clusters can be seen.

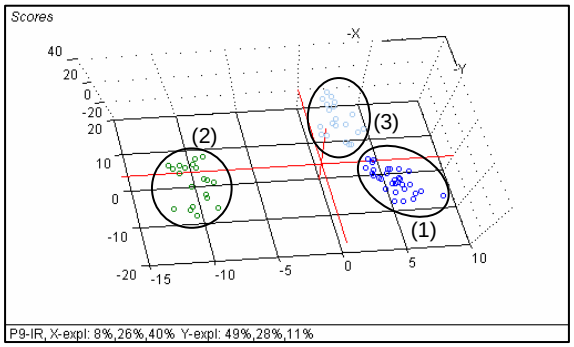


Figure 3.6: Score plot for Clairette Blanche sample set 2005A and 2005B using only FTIR data..
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

3.3.1.1.5 Discrimination of wines for combined data for 2005 and 2006

The score plot in Fig. 3.7 was produced using only FTIR data. Two clusters are evident, the one is wines fermented with WE372 (added in 2006) and the other contains wines fermented with yeast strains VIN13, VIN13-EXS and VIN13-PPK. The score plot of wines fermented with yeast strains VIN13, VIN13-EXS and VIN13-PPK (not shown), after leaving out WE372, show no clustering of fermented wines.

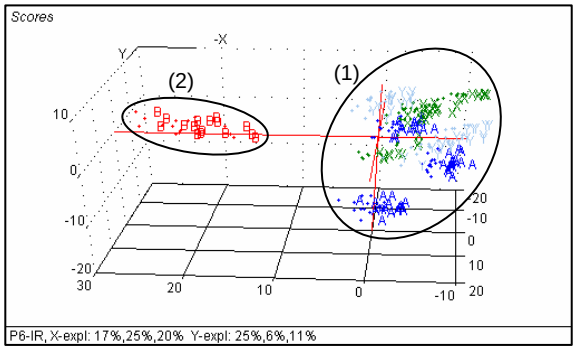


Figure 3.7: Score plot for Clairette Blanche sample set from 2005 and 2006 using only FTIR data.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1); o - VIN13-PPK (1)

Table 3.4: Results summary for Clairette Blanche

Set	Validation	Data used	Yeast used	Result	Comment	Type	Discrim	Fig.
2005A	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	tight clusters, far apart, 3 yeasts	score	yes	3
2005B	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	tight clusters, far apart, 3 yeasts	score	yes	4
2005B	leverage	GC	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	3 yeasts	score	yes	5
2006A	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	4 clusters	tight clusters, 4 yeasts	score	yes	6
2006A	leverage	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	4 clusters	tight clusters, 4 yeasts	score	yes	7
2005A, 2005B	leverage	FTIR	VIN13,VIN13-EXS,VIN13-PPK	3 clusters	tight clusters, far apart, 3 yeasts	score	yes	8
2005, 2006	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	2 clusters	WE372 vs. rest	score	yes	9

3.3.1.2 Pinotage

Fig. 3.8 represents a summary of the experimental layout for the wines made with the cultivar Pinotage. It covers the two vintages, 2005 and 2006, the independent sample sets drawn from the wines and the yeast used to ferment the musts with. During the 2005 vintage no fermentations were done to compile a test set and therefore leverage

correction was used as validation method in the PLS2 models. For the 2006 vintage, a separate set of fermentations was done that could be used as a test set during validation. This test set used the same grapes, but was de-stemmed and crushed two days after the first set of fermentations, using the same yeast but prepared independently. This second set of fermentations was used for test set validation, unless otherwise specified. A summary of results are presented in Table 3.5 at the end of the section.

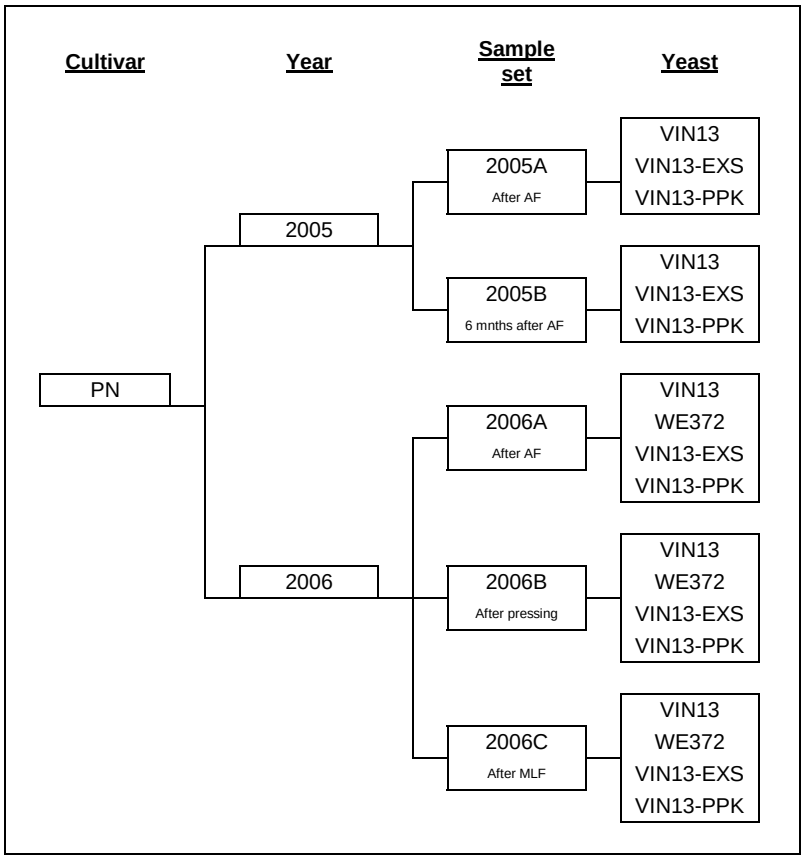


Figure 3.8: Experimental layout for wines made with cultivar Pinotage.

3.3.1.2.1 Discrimination of wines in sample set 2005A

The sample set was analysed using FTIR. Figure 3.9 shows the resulting score plot. Three clusters of objects that represent wines fermented with the three yeasts strains can be distinguished.

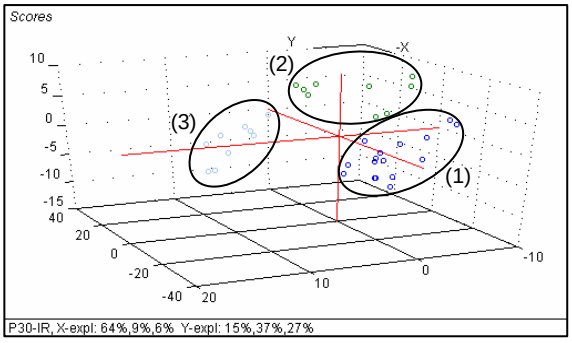


Figure 3.9: Score plot for Pinotage sample set 2005A using only FTIR data.
o - VIN13 (1); o - VIN13-EXS (2); o - VIN13-PPK (3)

3.3.1.2.2 Discrimination of wines in sample set 2005B

The sample set was analysed using FTIR spectroscopy and gas chromatography.

Using only the FTIR data, the resulting score plot is shown in Fig. 3.10a. Three tight clusters, spread far apart, representing the wines fermented with the three yeast strains used.

Fig. 3.10b shows the score plot of the data set where only GC data was used. Three clusters, spread far apart can be seen, although the clusters are not as tight.

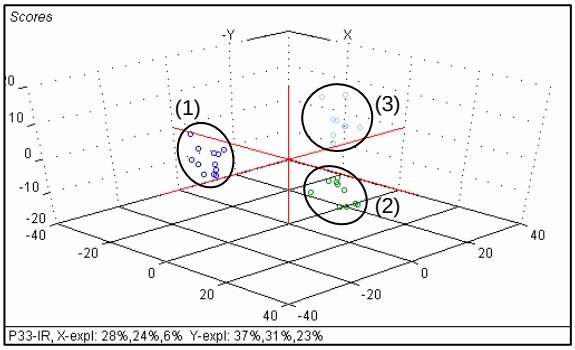


Figure 3.10a: Score plot for Pinotage sample set 2005B using only FTIR data.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

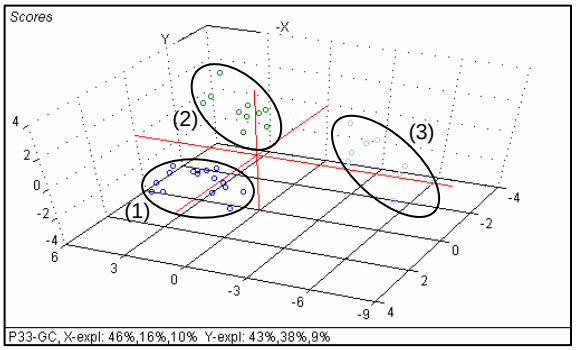


Figure 3.10b: Score plot for Pinotage sample set 2005B using only GC data.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (3)

3.3.1.2.3 Discrimination of wines in sample set 2006A

Even though a test set was available for 2006 fermentations, only the first set of fermentation in the 2006A sample set was analysed with FTIR. Leverage correction was used as method of validation.

Fig. 3.11 shows the score plot for the FTIR data. Four clusters can be observed, spread far apart, representing the wines fermented with the four different yeasts.

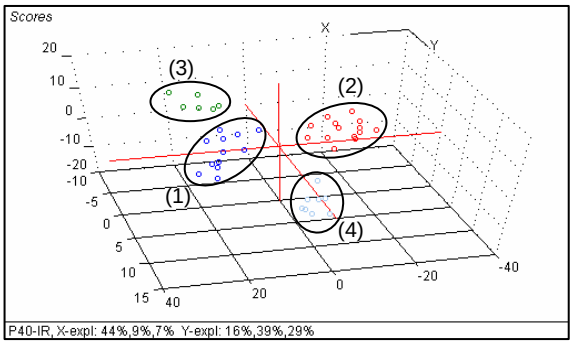


Figure 3.11: Score plot for Pinotage sample set 2006A using only FTIR data.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (3); **o** - VIN13-PPK (4)

3.3.1.2.4 Discrimination of wines in sample set 2006B

As can be seen from the score plot in Fig. 3.12a using FTIR data only there are two clusters, the one is wines fermented with WE372 and the other is the wines fermented with yeast strains VIN13, VIN13-EXS and VIN13-PPK.

Fig. 3.12b show the Predicted vs. Measured plot for wines fermented with WE372 against the other three wines. Discrimination can be observed by looking at the vertical gap between the two groups. The Predicted vs. Measured plots (not shown here) for the other wines show no discrimination.

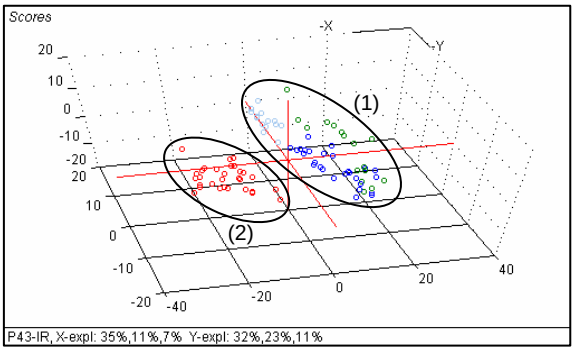


Figure 3.12a: Score plot for Pinotage sample set 2006B using only FTIR data. Test set validation. **o** - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (1); **o** - VIN13-PPK (1)

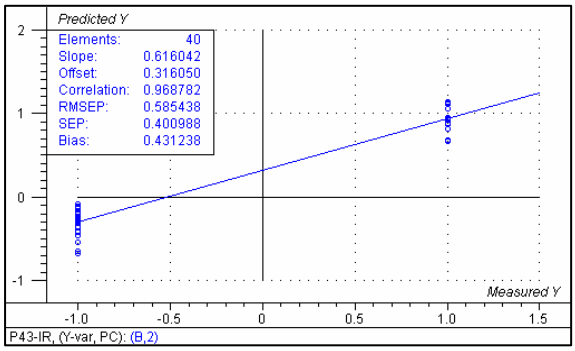


Figure 3.12b: Predicted vs. Measured plot for the Pinotage sample set 2006B using only FTIR data for prediction of WE372. Test set validation.

To get some more insight in the possible discrimination of VIN13, VIN13-EXS and VIN13-PPK a new PLS2-Discrim was done with the removal of all fermentations done with WE372. Fig. 3.13a shows the resulting score plot. Two clusters can be seen, the one represent wines fermented with VIN13-PPK, the other show the wines fermented with VIN13 and VIN13-EXS. Fig. 3.13b shows the Predicted vs. Measured plot for VIN13-PPK against the other two yeasts. No discrimination can be observed even though a vertical gap is present; this is due to the predicted values that are positive in value for some of the wines fermented with VIN13 and VIN13-EXS. The Predicted vs. Measured plots (not shown here) for the other wines also show no discrimination.

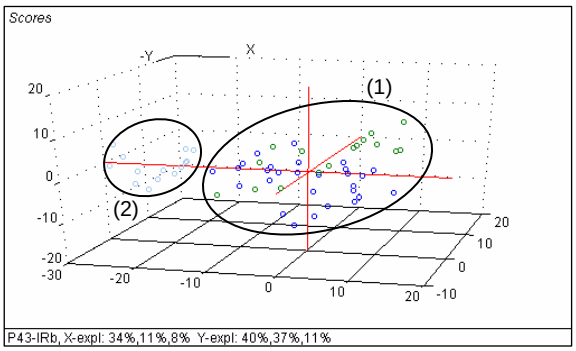


Figure 3.13a: Score plot for Pinotage sample set 2006B using only FTIR data. Test set validation. WE372 samples removed from sample set 2006B. **o** - VIN13 (1); **o** - VIN13-EXS (1); **o** - VIN13-PPK (2)

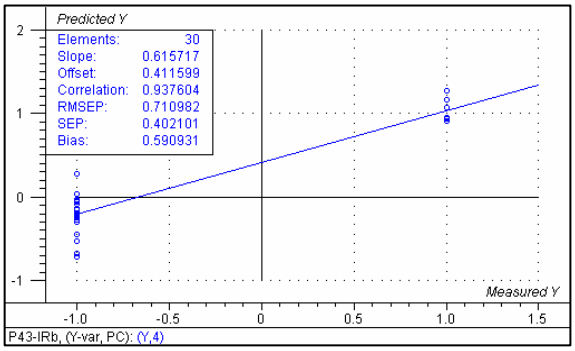


Figure 3.13b: Predicted vs. Measured plot for the Pinotage sample set 2006B using only FTIR data for prediction of VIN13-PPK after all samples produced from WE372 was removed. Test set validation.

3.3.1.2.5 Discrimination of wines in sample set 2006C

The sample set was analysed using FTIR spectroscopy and GC.

Fig. 3.14a shows the score plot of the data set where only FTIR data were used. There are two clusters representing wines fermented with WE372 and the other with wines fermented with yeast VIN13, VIN13-EXS and VIN13-PPK. VIN13-PPK can almost be seen as occupying a separate cluster with only a small overlap with VIN13 and VIN13-EXS.

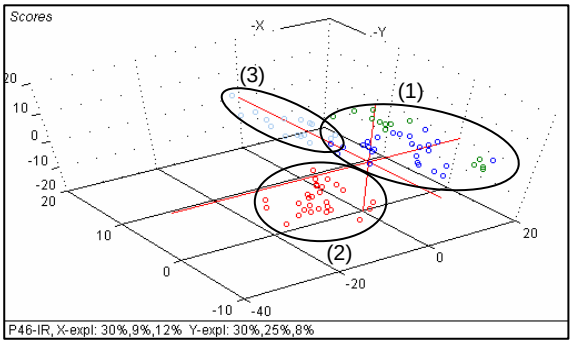


Figure 3.14a: Score plot for Pinotage sample set 2006C using only FTIR data. Test set validation.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (1); **o** - VIN13-PPK (3)

There is good discrimination on the Predicted vs. Measured plot for WE372 (Fig. 3.14b) and VIN13-PPK (Fig. 3.14c) using the data from all the wines.

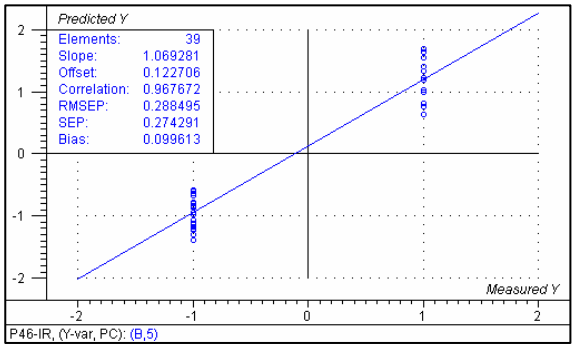


Figure 3.14b: Predicted vs. Measured plot for the Pinotage sample set 2006C using only FTIR data for prediction of WE372. Test set validation.

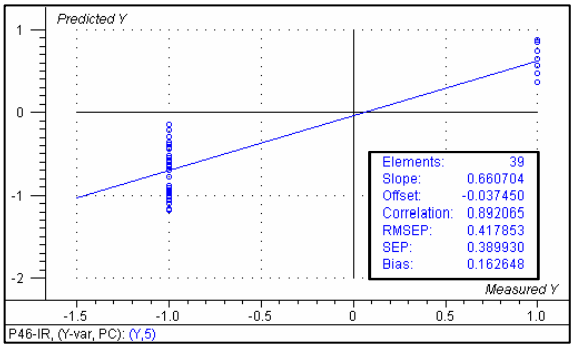


Figure 3.14c: Predicted vs. Measured plot for the Pinotage sample set 2006C using only FTIR data for prediction of VIN13-PPK. Test set validation.

Fig. 3.15a shows the score plot of the data set where the GC data were used. There are three clusters. One cluster representing wines fermented with WE372, one representing wines fermented with VIN13-PPK and the other representing wines fermented with VIN13 and VIN13-EXS. This discrimination can also be seen in the Predicted vs. Measured plots for wines fermented with WE372 (Fig. 3.15b) and VIN13-PPK (Fig. 3.15c).

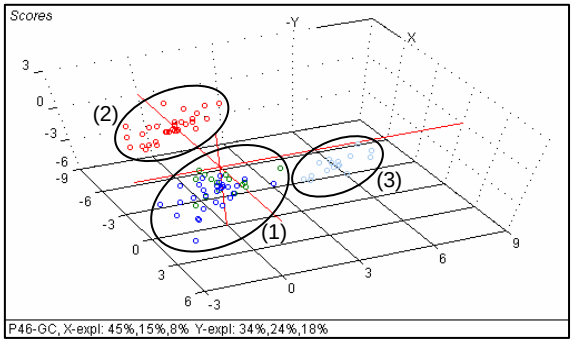


Figure 3.15a: Score plot for Pinotage sample set 2006C using only GC data. Test set validation.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (1); **o** - VIN13-PPK (3)

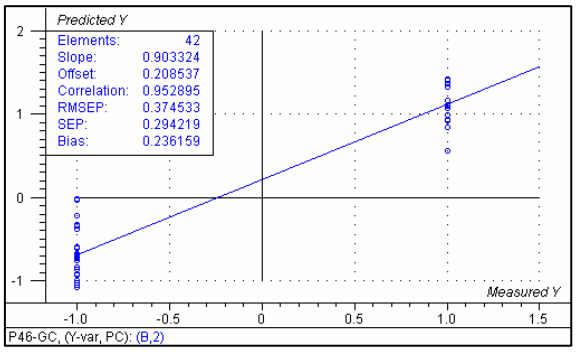


Figure 3.15b: Predicted vs. Measured plot for the Pinotage sample set 2006C using only GC data for prediction of WE372. Test set validation.

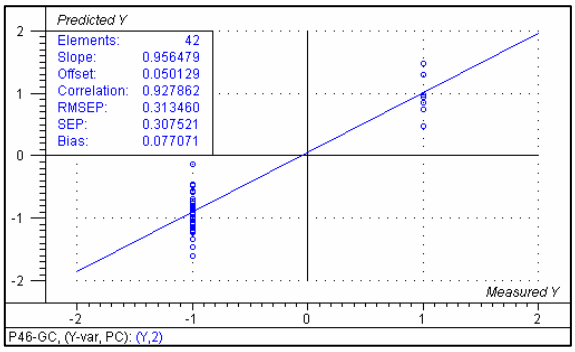


Figure 3.15c: Predicted vs. Measured plot for the Pinotage sample set 2006C using only GC data for prediction of VIN13-PPK. Test set validation.

3.3.1.2.6 Discrimination of wines for combined 2005 data

Fig. 3.16 shows the score plot for the combination of 2005 sample sets using FTIR data. Three clusters are observed but with some overlap. The Prediction vs. Measured plots (not shown) does not show any discrimination.

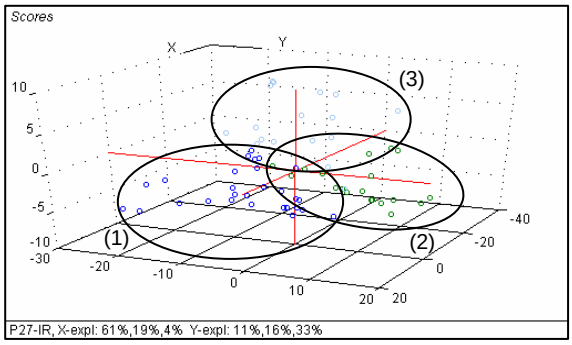


Figure 3.16: Score plot for Pinotage combined 2005 sample sets using only FTIR data. Test set validation.
o - VIN13 (1); o - VIN13-EXS (2); o - VIN13-PPK (3)

3.3.1.2.7 Discrimination of wines for combined 2006 data

Fig. 3.17a shows the score plot for the combined 2006 sample sets using FTIR data. There are three clusters, one of wines fermented with WE372, one of wines fermented with VIN13-PPK and one of wines fermented with VIN13 and VIN13-EXS. There is some overlap between all of these clusters and this is confirmed by the Predicted vs. Measured plots (not shown) with no gap between groups. The Predicted vs. Measured plot is shown in Fig. 3.17b of wines fermented with WE372. There is

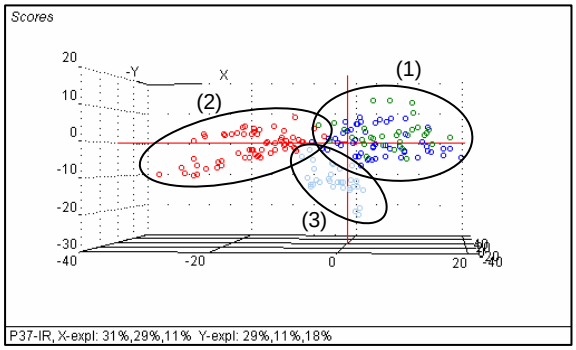


Figure 3.17a: Score plot for Pinotage for all 2006 sample sets using only FTIR data. Test set validation.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (3)

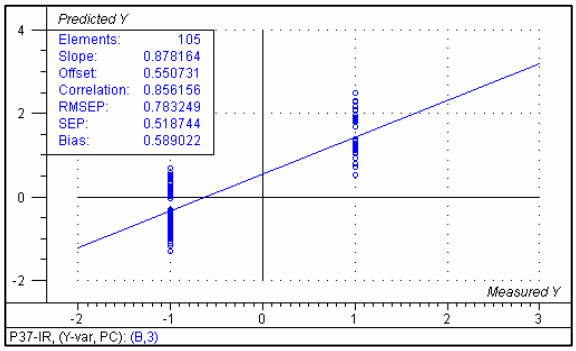


Figure 3.17b: Predicted vs. Measured plot for all Pinotage 2006 sample sets using only FTIR data for prediction of wines fermented with WE372. Test set validation. Overlap between objects of WE372 and the rest of the objects can be seen.

significant overlap and it is clear that no discrimination can be found between wines fermented with WE372 and the rest.

3.3.1.2.8 Discrimination of wines for combined data for 2005 and 2006

Fig. 3.18a shows the score plot of the combined sample sets of 2005B and 2006C using FTIR data and leverage correction as validation method. Two clusters can be seen, one with wines fermented with WE372 (only used in 2006), the second of wines fermented with VIN13-PPK but overlapping with wines fermented with VIN13 and VIN13-EXS.

Fig. 3.18b shows the score plot of the combined sample sets of 2005B and 2006C using GC data and leverage correction as validation method. Three clusters can be seen, the first of wines fermented with WE372 (only used in 2006), the second of wines fermented with VIN13-PPK and the other a combination of wines fermented with VIN13 and VIN13-EXS.

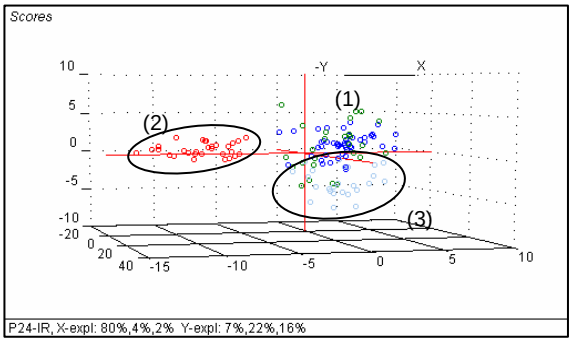


Figure 3.18a: Score plot for Pinotage sample sets for 2005 and 2006 using only FTIR data. Leverage correction.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1); o - VIN13-PPK (3)

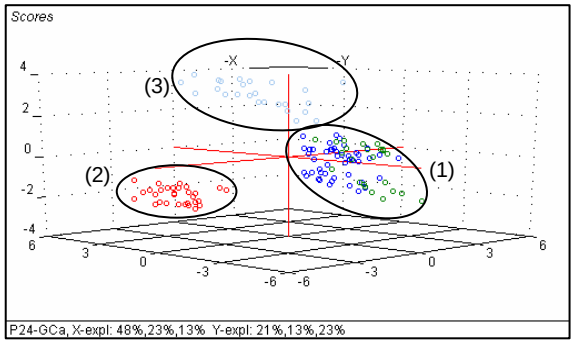


Figure 3.18b: Score plot for Pinotage sample sets for 2005 and 2006 using only GC data. Leverage correction.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1); o - VIN13-PPK (3)

Table 3.5: Results summary for Pinotage

Set	Validation	Data used	Yeast used	Result	Comment	Type	Discrim	Fig.
2005A	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	3 yeast	score	yes	11
2005B	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	tight clusters, far apart, 3 yeast	score	yes	12
2005B	leverage	GC	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	clusters spread out, far apart	score	yes	13
2006A	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	4 clusters	spread far apart, 4 yeast	score	yes	14
2006B	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	2 clusters	WE372 vs. (VIN13, VIN13-EXS,VIN13-PPK)	score	yes	15

Table 3.5: Results summary for Pinotage (continued)

Set	Validation	Data used	Yeast used	Result	Comment	Type	Discrim	Fig.
2006B	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	WE372 vs. rest		P v M	yes	16
2006B	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	2 clusters	VIN13-PPK vs. AX, only some overlap between VIN13 and VIN13-EXS	score	yes	17
2006B	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	VIN13-PPK vs. rest		P v M	no	18
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	2 clusters	WE372 vs. (VIN13, VIN13-EXS,VIN13-PPK), VIN13-PPK small overlap with (VIN13, VIN13-EXS)	score	yes	19
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	WE372 vs. rest		P v M	yes	20
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	VIN13-PPK vs. rest		P v M	yes	21
2006C	test	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	3 clusters	WE372, VIN13-PPK, (VIN13, VIN13-EXS)	score	yes	22
2006C	test	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	WE372 vs. rest		P v M	yes	23
2006C	test	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	VIN13-PPK vs. rest		P v M	yes	24
2005A, 2005B	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	Cluster overlap each other	score	no	25
2006A, 2006B, 2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	3 clusters	Cluster WE372, cluster VIN13-PPK, cluster (VIN13, VIN13-EXS)with overlap between all	score	no	26
2006A, 2006B, 2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	WE372 vs. rest	no discrimination	P v M	no	27
2005B, 2006C	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	2 clusters	WE372, VIN13-PPK overlap (VIN13, VIN13-EXS)	score	yes	28
2005, 2006	leverage	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	3 clusters	WE372, VIN13-PPK, (VIN13, VIN13-EXS)	score	yes	29

3.3.1.3 Cabernet Sauvignon

Fig. 3.19 shows the experimental layout for the wines made with Cabernet Sauvignon. It covers the two vintages, 2005 and 2006, the independent sample sets drawn from the wines and the yeast used to ferment the musts with.

During the 2005 vintage no fermentations were done to compile a test set and therefore leverage correction was used as validation method in the PLS2 models. For the 2006 vintage, a separate set of fermentations was done that could be used as a test set during validation. This test set used the same grapes, but was de-stemmed and crushed two days after the first set of fermentations, using the same yeast but prepared independently. This second set of fermentations was used for test set validation, unless otherwise specified.

A summary of results are presented in Table 3.6 at the end of the section.

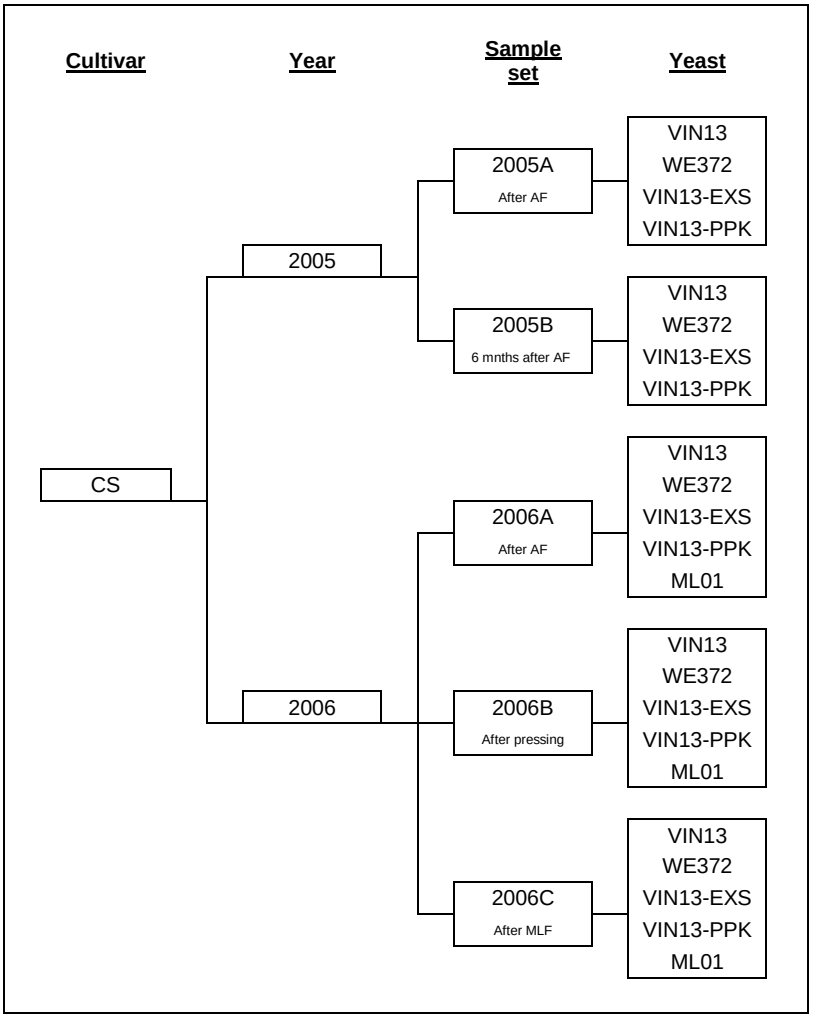


Figure 3.19: Experimental layout for wines made with varietal Cabernet Sauvignon.

3.3.1.3.1 Discrimination of wines in sample set 2005A

Fig. 13.20a shows the score plot for sample set 2005A using FTIR data and leverage correction. A cluster containing wines fermented with WE372 can be seen while there is some pattern in the rest of the objects. There is a group of wines fermented with VIN13 on the positive side of the Z axis. Another group containing wines fermented with VIN13-EXS and VIN13-PPK is on the negative side of the Z axis. Once the wines fermented with WE372 are removed form the sample set we can see three clusters in Fig. 13.20b. The three clusters represent wines fermented with VIN13, wines fermented with VIN13-EXS with an overlap with wines fermented with VIN13-PPK.

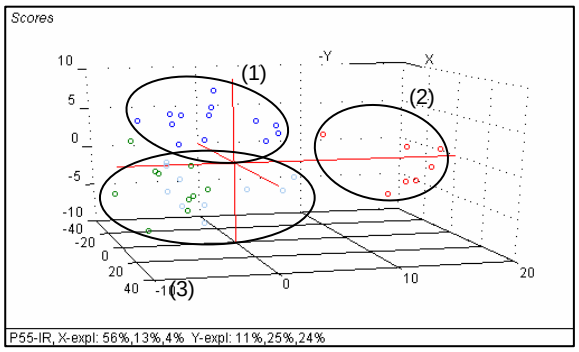


Figure 13.20a: Score plot for Cabernet Sauvignon for sample set 2005A using only FTIR data.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (3);
o - VIN13-PPK (3)

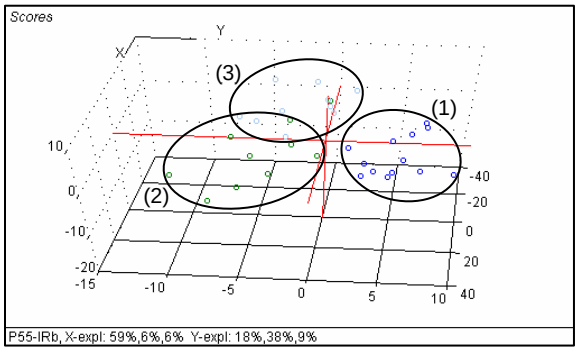


Figure 13.20b: Score plot for Cabernet Sauvignon for sample set 2005A using only FTIR data. WE372 samples removed from sample set 2005B.
o - VIN13 (1); o - VIN13-EXS (2); o - VIN13-PPK (3)

3.3.1.3.2 Discrimination of wines in sample set 2005B

Fig. 13.21a shows the score plot for sample set 2005B using FTIR data and leverage correction. A cluster containing wines fermented with WE372 can be seen while there is no pattern in the rest of the objects. Once the wines fermented with WE372 are removed form the sample set we can see three clusters in Fig. 13.21b. The three clusters represent wines fermented with VIN13, wines fermented with VIN13-EXS and wines fermented with VIN13-PPK.

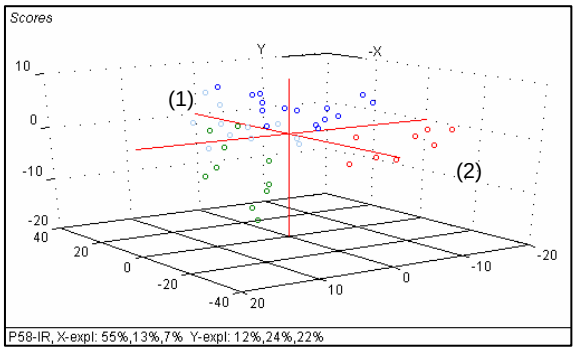


Figure 13.21a: Score plot for Cabernet Sauvignon for sample set 2005B using only FTIR data.
o – VIN13 (1); o – WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (1)

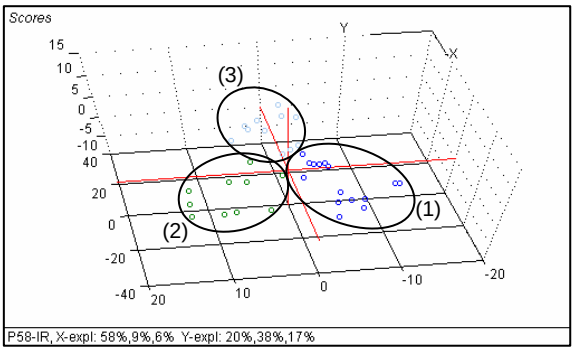


Figure 13.21b: Score plot for Cabernet Sauvignon for sample set 2005B using only FTIR data. Yeast B samples removed from sample set 2005B.
o – VIN13 (1); o - VIN13-EXS (2); o - VIN13-PPK (3)

Fig. 13.22 shows the score plot for sample set 2005B using GC data and leverage correction. Four clusters representing wines fermented with VIN13, WE372, VIN13-EXS and VIN13-PPK can be seen.

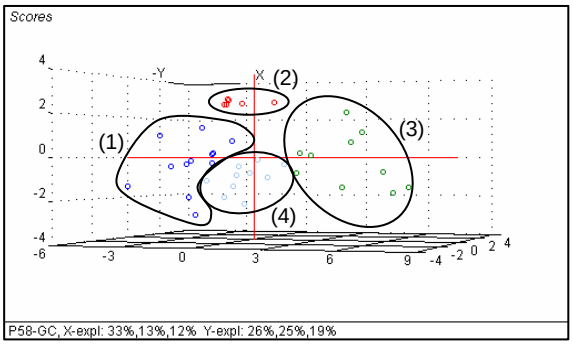


Figure 13.22: Score plot for Cabernet Sauvignon for sample set 2005B using only GC data.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (3);
o - VIN13-PPK (4)

3.3.1.3.3 Discrimination of wines in sample set 2006A

Fig. 13.23a shows the score plot for sample set 2006A using FTIR data and test set validation. There are three clusters spread far apart in score space. The three clusters represent wines fermented with WE372, ML01 and a cluster containing wines fermented with VIN13, VIN13-EXS and VIN13-PPK.

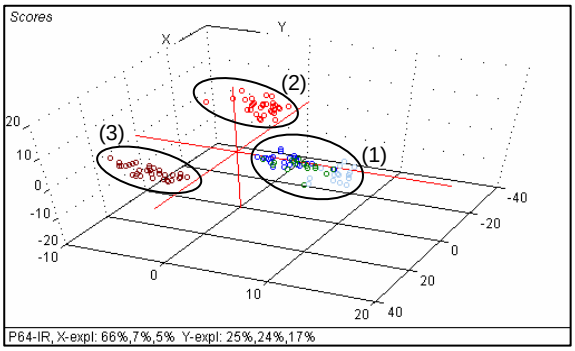


Figure 13.23a: Score plot for Cabernet Sauvignon for sample set 2006A using only FTIR data. Test set validation.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (1);
o - VIN13-PPK (1); **o** - ML01 (3)

Fig. 13.23b shows the Predicted vs. Measured plot to illustrate the discrimination between wines fermented with WE372 and the other. The Predicted vs. Measure plot in Fig. 13.23c shows the discrimination between wines fermented with ML01 against the others. Both have good discrimination with a wide gap between groups.

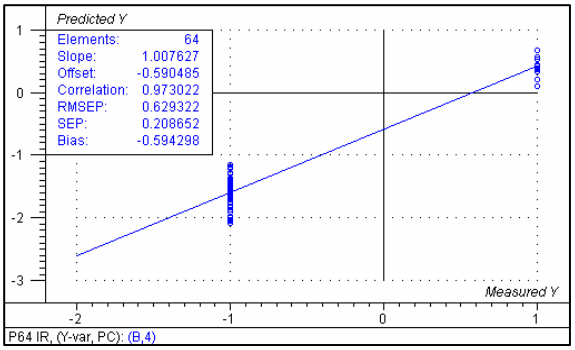


Figure 13.23b: Predicted vs. Measured plot for the Cabernet Sauvignon 2006A sample set using only FTIR data for prediction of wines fermented with WE372. Test set validation.

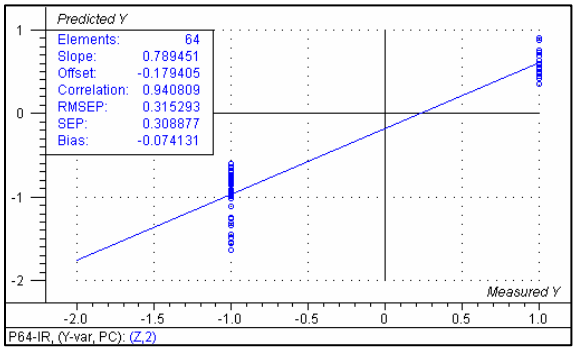


Figure 13.23c: Predicted vs. Measured plot for the Cabernet Sauvignon 2006A sample set using only FTIR data for prediction of wines fermented with ML01. Test set validation.

After removing wines fermented with WE372 and ML01 it can be seen in Fig. 13.23d that wines fermented with VIN13 and VIN13-EXS form one cluster and wines fermented with VIN13-PPK form another.

The Prediction vs. Measured plot in Fig. 13.23e shows that there is no discrimination between wines fermented with VIN13-PPK and those fermented with VIN13 and VIN13-EXS.

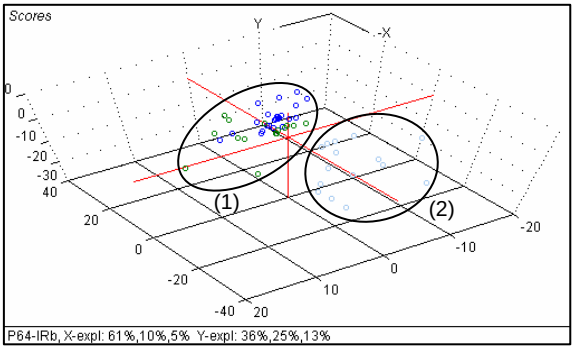


Figure 13.23d: Score plot for Cabernet Sauvignon for sample set 2006A using only FTIR data. WE372 and ML01 samples removed from sample set 2006A. Test set validation.
o - VIN13 (1); **o** - VIN13-EXS (1); **o** - VIN13-PPK (2)

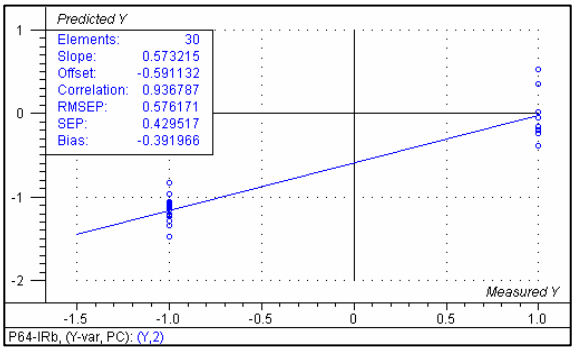


Figure 13.23e: Predicted vs. Measured plot for the Cabernet Sauvignon 2006A sample set using only FTIR data for prediction of wines fermented with VIN13-PPK. WE372 and ML01 samples removed from sample set 2006A. Test set validation.

3.3.1.3.4 Discrimination of wines in sample set 2006B

The score plot for sample set 2006B using FTIR data and test set validation is shown in Fig. 13.24a. Three clusters can be seen, representing wines fermented with WE372, ML01 and the other cluster wines fermented with VIN13, VIN13-EXS and VIN13-PPK.

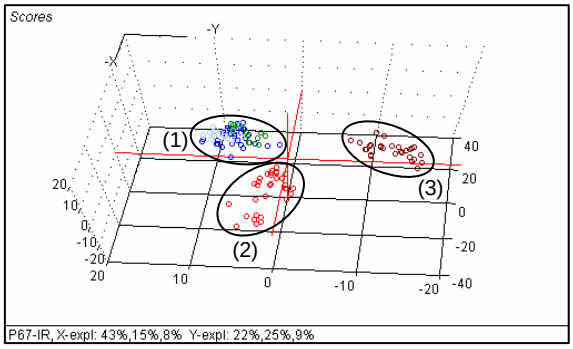


Figure 13.24a: Score plot for Cabernet Sauvignon for sample set 2006B using only FTIR data. Test set validation.
o - VIN13 (1); **o** - WE372 (2); **o** - VIN13-EXS (1); **o** - VIN13-PPK (1); **o** - ML01 (3)

The Predicted vs. Measured plots of wines fermented with WE372 and ML01 are shown in Figs. 13.24b and 13.24c. Good discrimination is obtained between wines fermented with WE372 and the rest. Wines fermented with ML01 also show good discrimination from the rest of the wines.

When the wines fermented with WE372 and ML01 were removed from the sample set, no discrimination could be seen between wines fermented with VIN13, VIN13-EXS and VIN13-PPK (plots not shown).

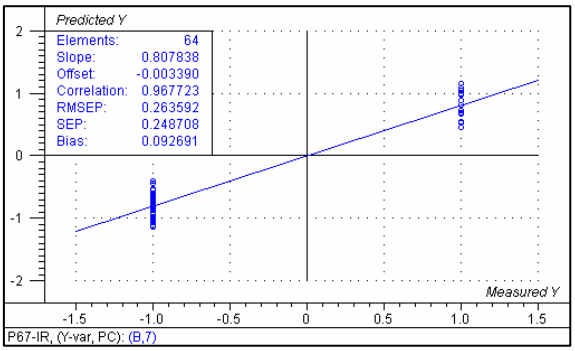


Figure 13.24b: Predicted vs. Measured plot for the Cabernet Sauvignon 2006B sample set using only FTIR data for prediction of wines fermented with WE372. Test set validation.

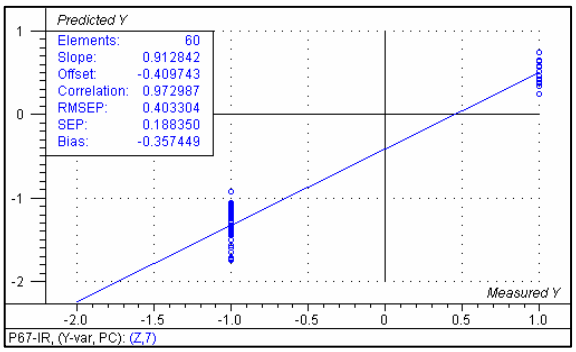


Figure 13.24c: Predicted vs. Measured plot for the Cabernet Sauvignon 2006B sample set using only FTIR data for prediction of wines fermented with ML01. Test set validation.

Fig. 13.24d shows the score plot for sample set 2006B using GC data and test set validation. The clusters are not as clearly defined as for the FTIR data. The three clusters for wines fermented with WE372, wines fermented with ML01 and wines fermented with VIN13, VIN13-EXS and VIN13-PPK can still be seen, even though there is some overlap between clusters.

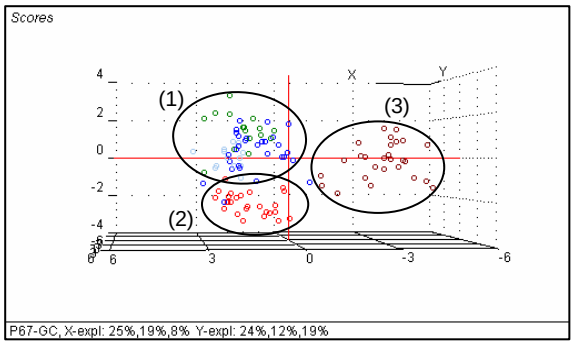


Figure 13.24d: Score plot for Cabernet Sauvignon for sample set 2006B using only GC data. Test set validation.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (1); o - ML01 (3)

The Predicted vs. Measured plots for the wines fermented with WE372 and ML01 (not shown) indicate that there will be overlap between the different clusters.

3.3.1.3.5 Discrimination of wines in sample set 2006C

Fig. 13.25a shows the score plot for sample set 2006C using FTIR data and test set validation. There are three tight clusters spread far apart, representing wines fermented with WE372, wines fermented with ML01 and wines fermented with VIN13, VIN13-EXS and VIN13-PPK. In the Predicted vs. Measured plots in Figs. 13.25b and 13.25c, discrimination is achieved with large vertical gaps between the different wines fermented with WE372 and ML01 and the other wines.

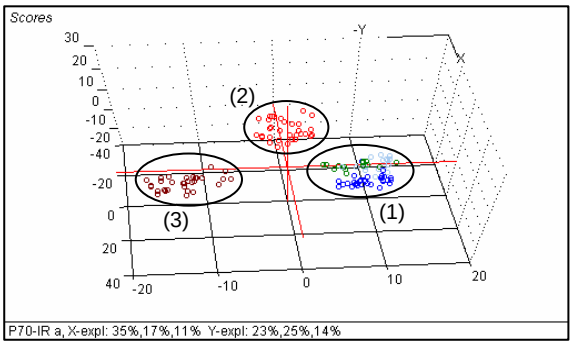


Figure 13.25a: Score plot for Cabernet Sauvignon for sample set 2006C using only FTIR data. Test set validation.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (1); o - ML01 (3)

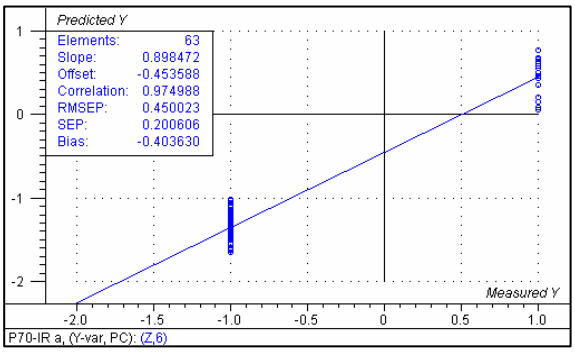


Figure 13.25b: Predicted vs. Measured plot for the Cabernet Sauvignon 2006C sample set using only FTIR data for prediction of wines fermented with ML01. Test set validation.

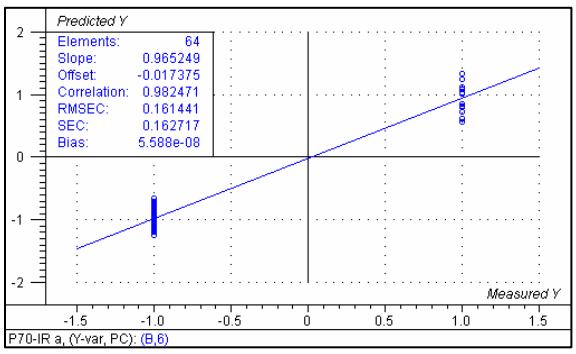


Figure 13.25c: Predicted vs. Measured plot for the Cabernet Sauvignon 2006C sample set using only FTIR data for prediction of wines fermented with WE372. Test set validation.

Fig. 13.25d shows the score plot after the wines fermented with WE372 and ML01 were removed from the sample set. There are two clusters, one containing wines fermented with VIN13 and another with wines fermented with VIN13-EXS and VIN13-PPK.

The Predicted vs. Measured plot in Fig. 13.25e shows the discrimination between wines fermented with VIN13 and the rest of the wines, with a wide vertical gap.

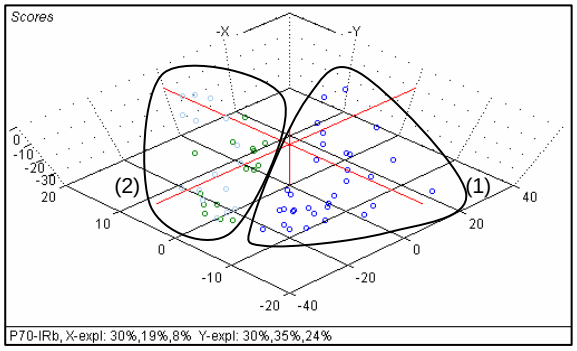


Figure 13.25d: Score plot for Cabernet Sauvignon for sample set 2006C using only FTIR data. Test set validation. WE372 and ML01 samples removed from sample set 2006C.
o - VIN13 (1); **o** - VIN13-EXS (2); **o** - VIN13-PPK (2)

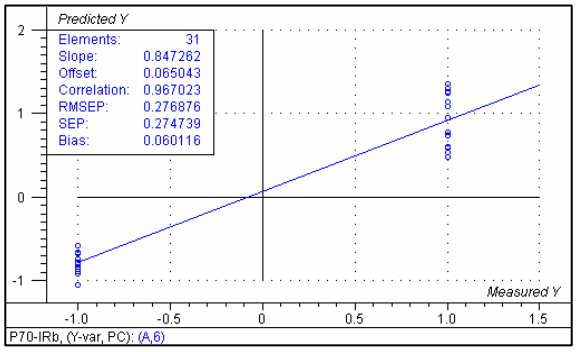


Figure 13.25e: Predicted vs. Measured plot for the Cabernet Sauvignon 2006C sample set using only FTIR data for prediction of wines fermented with VIN13. WE372 and ML01 samples removed from sample set 2006C. Test set validation.

3.3.1.3.6 Discrimination of wines for combined 2005 data

No clusters could be seen in the score plot of the combined 2005 data (score plot not shown).

3.3.1.3.7 Discrimination of wines for combined 2006 data

The score plot for the combined data for 2006 using FTIR data is shown in Fig. 13.26a. Three clusters can be seen, one with wines fermented with WE372, one with wines

fermented with ML01 and the other wines fermented with VIN13, VIN13-EXS and VIN13-PPK.

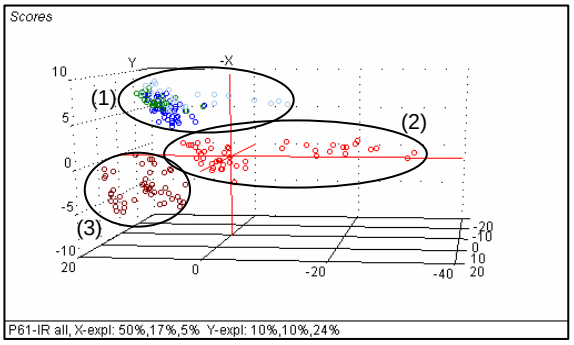


Figure 13.26a: Score plot for Cabernet Sauvignon for all 2006 sample sets using only FTIR data. Test set validation.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (1); o - ML01 (3)

Figs. 13.26b and 13.26c show the Predicted vs. Measured plots for wines fermented with WE372 and ML01, but shows that there is overlap between the clusters.

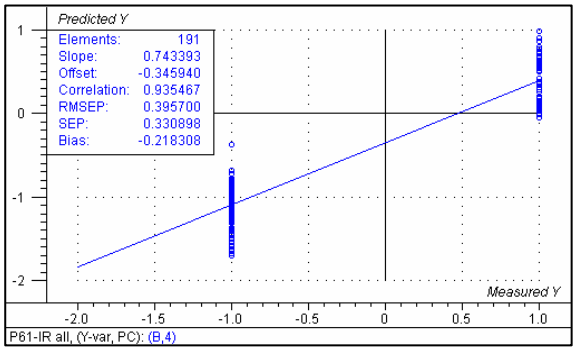


Figure 13.26b: Predicted vs. Measured plot for the Cabernet Sauvignon 2006 sample sets using only FTIR data for prediction of wines fermented with WE372. Test set validation.

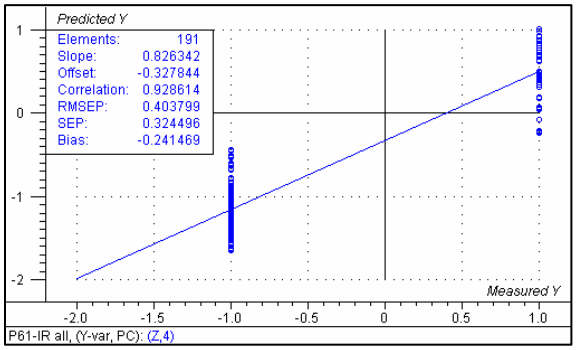


Figure 13.26c: Predicted vs. Measured plot for the Cabernet Sauvignon 2006 sample sets using only FTIR data for prediction of wines fermented with ML01. Test set validation.

3.3.1.3.8 Discrimination of wines for combined 2005 and 2006 data

With the combining of the FTIR data for the two vintages, sample set 2005B and 2006B and using leverage correction, three clusters can be seen in Fig. 13.27. These clusters represent wines fermented with WE372, wines fermented with ML01 and a combination of wines fermented with VIN13, VIN13-EXS and VIN13-PPK.

Using GC data only a cluster containing wines fermented with ML01 could be observed (score plot not shown).

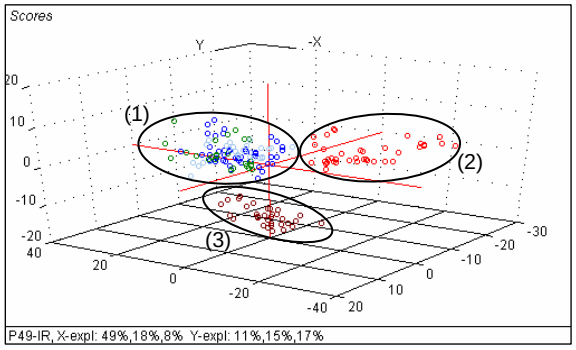


Figure 13.27: Score plot for combined sample set Cabernet Sauvignon 2005B and 2006B using only FTIR data. Leverage correction used.
o - VIN13 (1); o - WE372 (2); o - VIN13-EXS (1);
o - VIN13-PPK (1); o - ML01 (3)

Table 3.6: Results summary for Cabernet Sauvignon

Set	Validation	Data used	Yeast used	Result	Comment	Type	Discrim	Fig.
2005A	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	3 clusters	WE372, VIN13 overlap (VIN13-EXS, VIN13-PPK)	score	yes	31
2005A	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	VIN13, VIN13-EXS overlap VIN13-PPK	score	yes	32
2005B	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK	1 cluster	only WE372 cluster, no pattern in rest	score	yes	33
2005B	leverage	FTIR	VIN13, VIN13-EXS, VIN13-PPK	3 clusters	VIN13, VIN13-EXS and VIN13-PPK clusters, close together	score	yes	34
2005B	leverage	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK	4 clusters	VIN13, WE372, VIN13-EXS and VIN13-PPK clusters	score	yes	35
2006A	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	36
2006A	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	WE372 vs. rest		P v M	yes	37
2006A	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	ML01 vs. rest		P v M	yes	38
2006A	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	2 clusters	VIN13-PPK, (VIN13-EXS, VIN13-PPK)	score	yes	39
2006A	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	VIN13-PPK vs. rest		P v M	no	40
2006B	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	41
2006B	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	WE372 vs. rest	wide gap	P v M	yes	42
2006B	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	ML01 vs. rest	wide gap	P v M	yes	43
2006B	test	GC	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	44

Table 5: Results summary for Cabernet Sauvignon **(continued)**

Set	Validation	Data used	Yeast used	Result	Comment	Type	Discrim	Fig.
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	45
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	ML01 vs. rest	wide gap	P v M	yes	46
2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	WE372 vs. rest	wide gap	P v M	yes	47
2006C	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	2 clusters	VIN13, (VIN13-EXS, VIN13-PPK)	score	yes	48
2006C	test	FTIR	VIN13, VIN13-EXS, VIN13-PPK	VIN13 vs. rest	wide gap	P v M	yes	49
2006A, 2006B, 2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	50
2006A, 2006B, 2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	WE372 vs. rest		P v M	no	51
2006A, 2006B, 2006C	test	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	ML01 vs. rest		P v M	no	52
2005B, 2006B	leverage	FTIR	VIN13, WE372, VIN13-EXS, VIN13-PPK, ML01	3 clusters	WE372, ML01, (VIN13, VIN13-EXS, VIN13-PPK)	score	yes	53

3.3.2 EFFECT OF AGEING OF WINES ON DISCRIMINATION

In Fig. 13.28a, using GC data from the 2005B and 2006A sample sets for Clairette Blanche and leverage correction we can see two clusters representing wines samples at the two different times during the year.

Fig. 13.28b show the combined 2005A and 2005B Pinotage sample sets using leverage correction and FTIR data. Two clusters are formed representing wines from the 2005A sample set and the 2005B sample set.

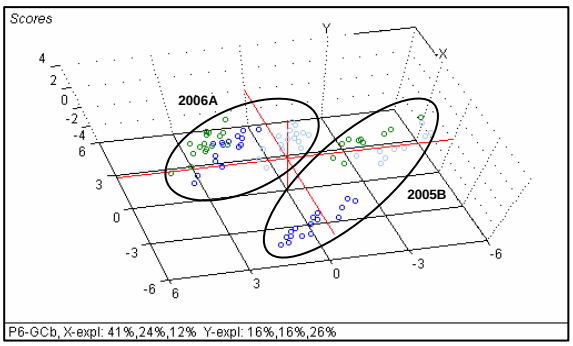


Figure 13.28a: Score plot for Clairette Blanche sample sets from 2005A and 2006B using only the GC data. Leverage correction. WE372 removed from sample set.
o - VIN13; o - VIN13-EXS; o - VIN13-PPK

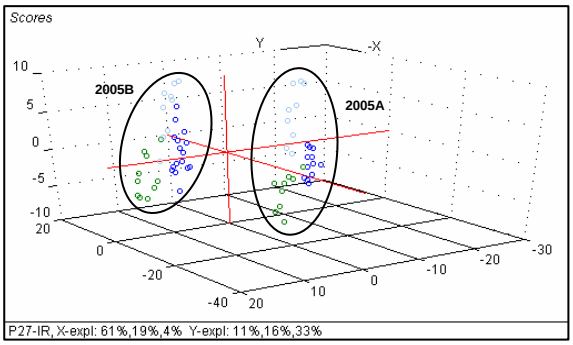


Figure 13.28b: Score plot for Pinotage combined 2005 sample sets using only FTIR data. Leverage correction. Clear discrimination between the two sample sets is observed.
o - VIN13; o - VIN13-EXS; o - VIN13-PPK

Using FTIR data for the combined 2006A, 2006B and 2006C Pinotage sample sets and test set validation, three clusters can be discriminated in Fig. 13.28c. The clusters represent wines from the 2006A, 2006B and 2006C sample sets.

Fig. 13.28d shows the score plot for the combined sample sets of 2005 using FTIR data and leverage correction. The two clusters represent the wines from the two samples sets.

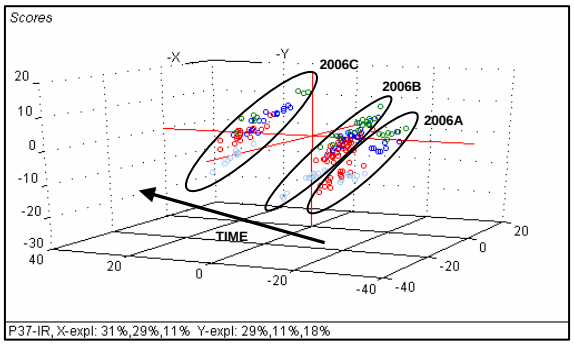


Figure 13.28c: Score plot for Pinotage for all 2006 sample sets using only FTIR data. Test set validation. Clustering between the different sample sets can be observed with very good discrimination between 2006B and 2006C.
o - VIN13; o - WE372; o - VIN13-EXS;
o - VIN13-PPK

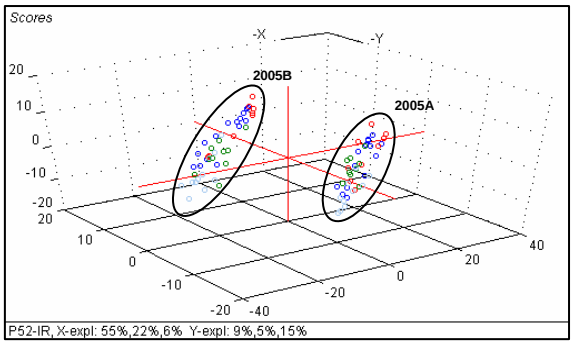


Figure 13.28d: Score plot for Cabernet Sauvignon for all 2005 sample sets using only FTIR data. Leverage correction. Clustering between the different sample sets can be observed with very good discrimination between 2005A and 2005B.
o - VIN13; o - WE372; o - VIN13-EXS;
o - VIN13-PPK

The combined data for 2006 using FTIR data and test set validation in the score plot shown in Fig. 13.28e does not show clustering for each sample set. There are two clusters for wines fermented with WE372, one cluster representing wines from sample set 2006A and the other wines from sample sets 2006B and 2006C. The pattern is repeated for wines fermented by WE372 and wines fermented by VIN13, VIN13-EXS and VIN13-PPK. It could be seen as 2006A wines to the top right and 2006B and 2006C towards the bottom left.

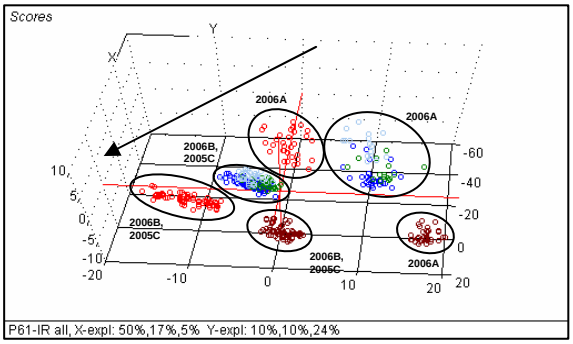


Figure 13.28e: Score plot for Cabernet Sauvignon for all 2006 sample sets using only FTIR data. Test set validation.
o - VIN13; **o** - WE372; **o** - VIN13-EXS;
o - VIN13-PPK; **o** - ML01

3.3.3 DISCRIMINATION BASED UPON NON-GM VS. GM YEAST STRAIN USED FOR FERMENTATION

A special case of discriminating wines based on the yeast used for fermentation would be to discriminate between wines fermented with non-genetically modified (non-GM) yeast and wines fermented with genetically modified (GM) yeast. In an attempt not to repeat all the plots covered in the preceding text only two cased will be reviewed.

3.3.3.1 PLS-DISCRIM data analysis of Clairette Blanche

Fig. 13.29a illustrates that discrimination is possible between GM-wines and non-GM-wines for the 2005 vintage model fermentations of Clairette Blanche. An excellent discrimination can be observed using either two or three PLS-components, based on FTIR spectroscopy only.

Fig. 13.29b demonstrates similar features for discrimionation between GM-wines and non-GM-wines for the case of the 2006 harvest. For this vintage a larger FTIR-complexity of the data (two sub-groups discernable both for GM- and for non-GM-wines) necessitates four PLS-components for the discrimination, which is otherwise fully comparable to that for for 2005 however. The sub-groups are clusters of the individual groups of yeast. VIN13 and WE372 form the LARGER non-GM wine group while VIN13-EXS and VIN13-PPK form the LARGER GM-wine group.

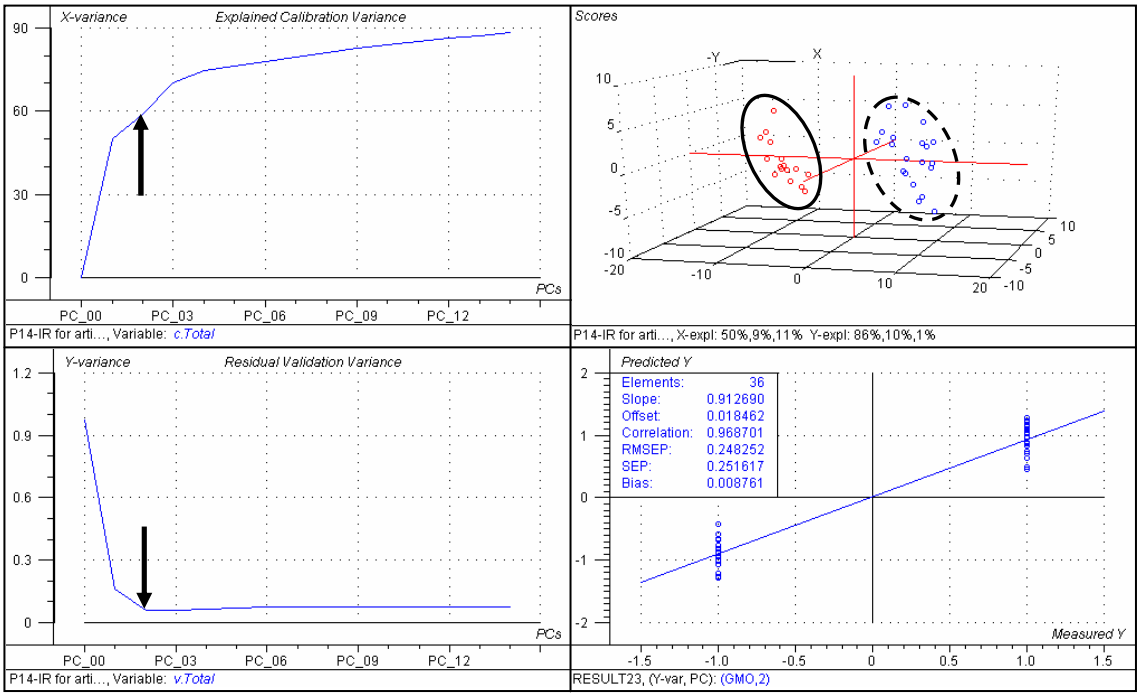


Fig. 13.29a. Clairette Blanche (2005 harvest). Top right panel: 3-D score plot. Bottom right panel: Predicted vs. measured validation (based on 2-segment cross validation). Clear discrimination gap between GM and non-GM predictions (+1 and -1 clusters respectively). Two-segmented cross-validation indicates 2 significant components. 94% Y-variable modelled, using 59% X-variance (FTIR data).
o – GM wines; **o** – non-GM wines

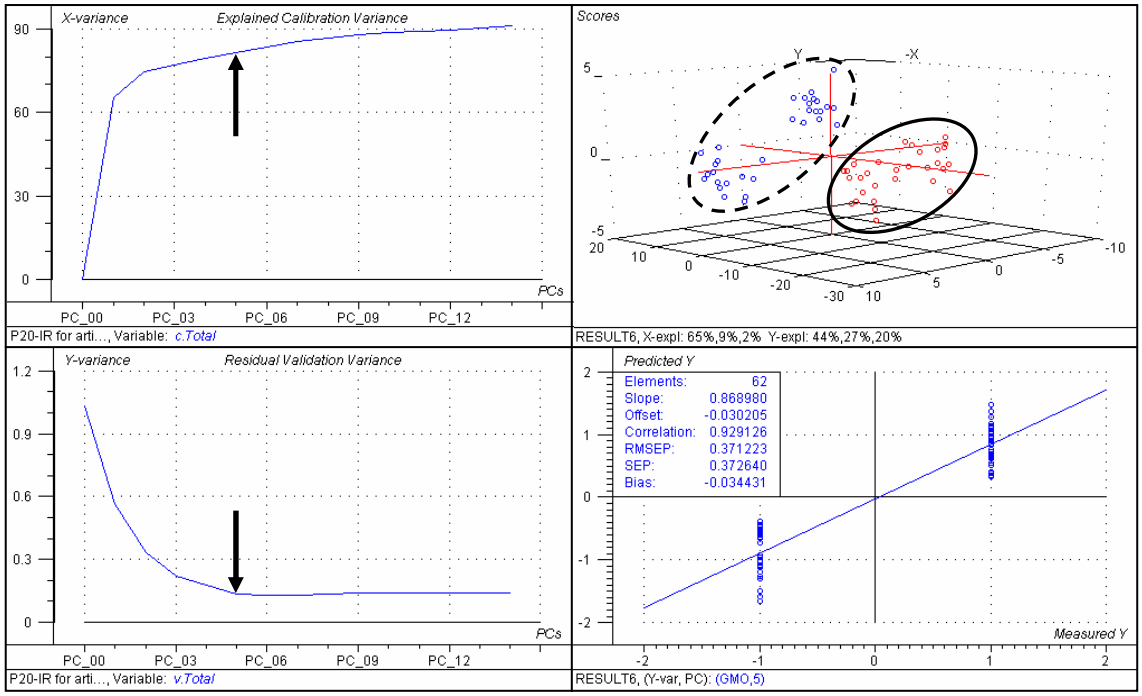


Fig. 13.29b. Clairette Blanche (2006 harvest). Top right panel: 3-D score plot. Bottom right panel: Predicted vs. measured validation (based on 2-segment cross validation). Clear discrimination gap between GM and non-GM predictions (+1 and -1 clusters respectively). Two-segmented cross-validation indicates 5 significant components. 87% Y-variable modelled, using 82% X-variance (FTIR data).
o – GM wines; **o** – non-GM wines

3.3.3.2 PLS-DISCRIM data analysis of Cabernet Sauvignon

As is evident from Fig. 13.30a, there is an excellent discrimination between GM-wines and those fermented by non-GM yeast for 2005 Cabernet Sauvignon. Four PLS-component are needed for this performance, all are statistically significant as per the validation (lower left panel).

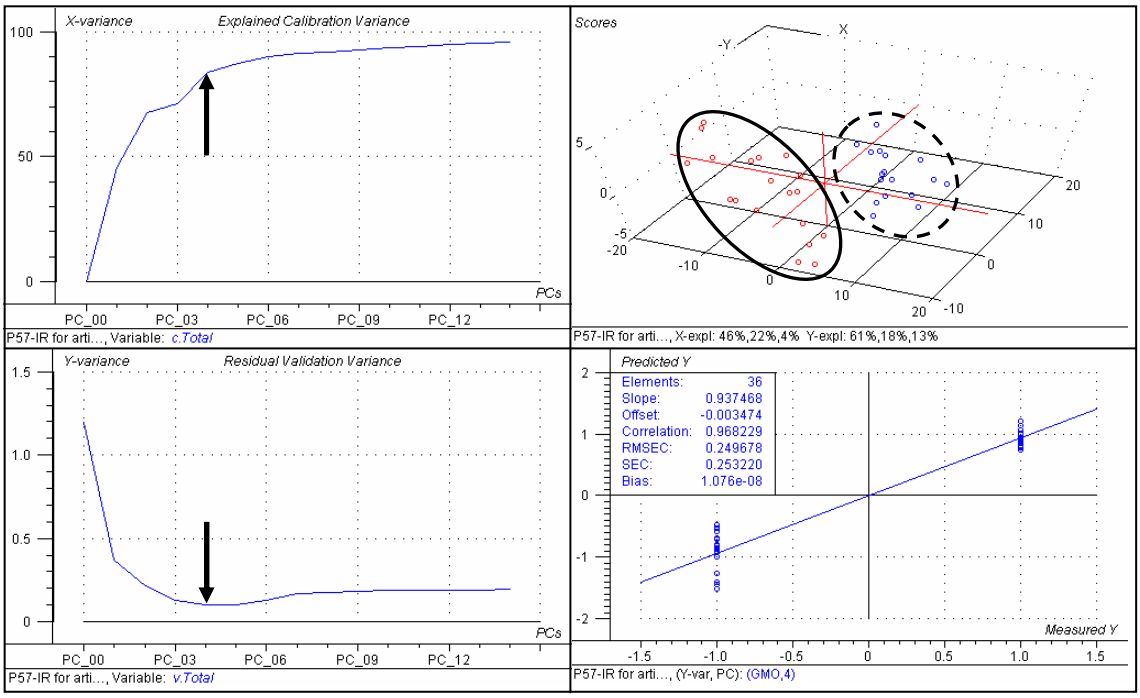


Fig. 13.30a. Cabernet Sauvignon (2005 harvest). Top right panel: 3-D score plot. Bottom right panel: Predicted vs. measured validation (based on 2-segment cross validation). An excellent discrimination gap between GM and non-GM predictions (+1 and -1 clusters respectively). Two-segmented cross-validation indicates 4 significant components. 87% Y-variable modelled, using 82% X-variance (FTIR data).
o – GM wines; **o** – non-GM wines

For 2006 the most stringent validation was used, independant test set validation. In Fig. 13.30b, a completely independent data set is used for the performance evaluation. Discrimination for Cabernet Sauvignon (2006) still comes through in a quite acceptable fashion. Prediction (of GM-discrimination in the case) cannot be tested in a more realistic sense that by this test set validation scenario (Esbensen, 2002).

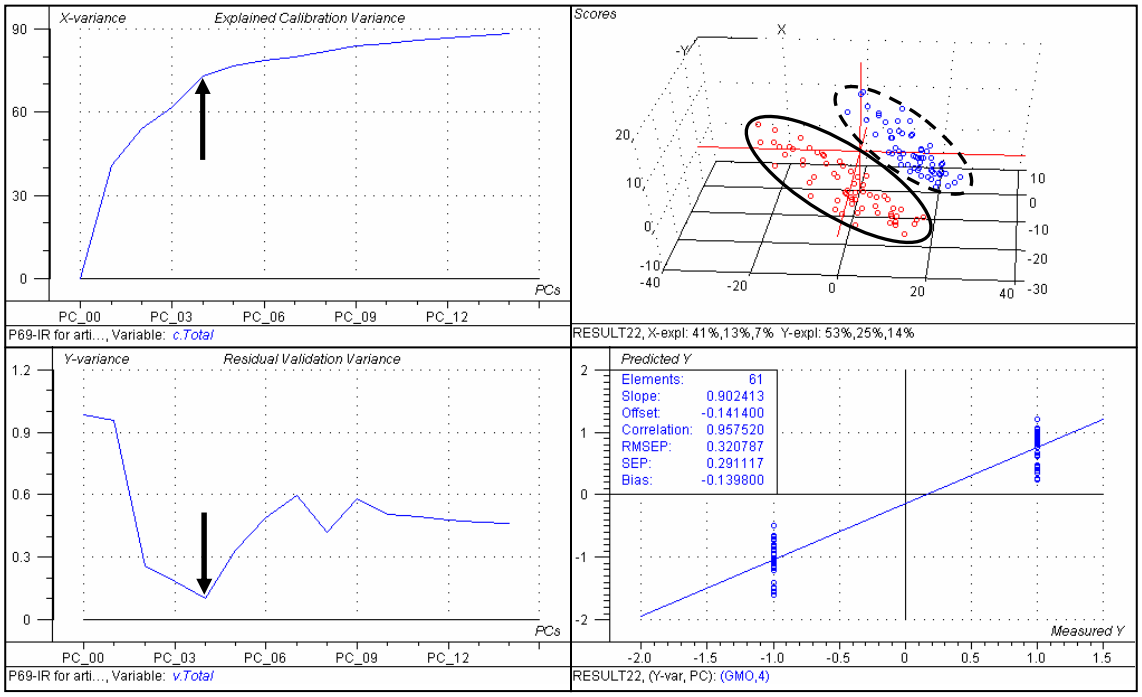


Fig. 13.30b. Cabernet Sauvignon (2006 harvest). Top right panel: 3-D score plot. Bottom right panel: Predicted vs. measured validation (N.B. test set validation). Again highly acceptable discrimination gap between GMO and non-GMO predictions (+1 and -1 clusters respectively). Validation indicates 4 significant components. 90% Y-variable modelled, using 73% X-variance (FTIR data).
o – GM wines; **o** – non-GM wines

3.4 DISCUSSION

3.4.1 DISCRIMINATION BETWEEN WINES FERMENTED WITH DIFFERENT YEAST STRAINS

When test set validation is used only objects from the calibration set is used to create the score plot and the test set objects are passively projected into this score space. When leverage correction is used and all objects are combined and no objects are kept out of the calibration, the score plot will be created using all the objects and therefore use a larger data set. When using leverage correction on combined calibration and test sets it has been shown (score plots not shown) that for some of the sample sets discrimination was possible between wines fermented with the closely related yeast strains.

3.4.1.1 Clairette Blanche

A tabulated summary of the results from wines fermented from Clairette Blanche grapes can be seen in Table 3.4.

For the FTIR data in the 2005A and in the 2005B sample sets good discrimination are achieved as seen by the tight clusters that are spread far apart in the score space.

Discrimination is possible using GC data in the 2005B sample set. Clusters are not as compact but no overlap is observed.

In the 2006A sample set using FTIR data good discrimination is achieved. Fermentations done with WE372 and VIN13-PPK is spread far from each other with the fermentations done with VIN13 and VIN13-EXS lies in between the fermentation made with WE372 and VIN13-PPK, but the fermentations with VIN13 and VIN13-EXS are still in separate clusters.

Using GC data for the 2006A sample set good discrimination is achieved with all the fermentation forming clusters that does not overlap with the others.

When combining the FTIR data from the two sample sets for 2005, good discrimination is achieved between the fermentations done with the three yeast strains. It seems possible that the fermentations done with VIN13 and VIN13-PPK are closer in terms of chemical composition than that of wines fermented with VIN13-EXS.

By combining the FTIR data of all the sample sets across the two vintages, 2005 and 2006, we can see if it is possible to discriminate between the fermentations done with the different yeasts also in successive years. This is important to create a successful model that could be used for prediction of the yeast used for a specific fermentation. The fermentations done with VIN13, VIN13-EXS and VIN13-PPK in 2005 and 2006, cluster together and seem to be more similar in chemical composition than those fermentations done with WE372 in 2006. The fact that wines fermented with VIN13, VIN13-EXS and VIN13-PPK cluster together for 2005 and 2006 could possibly indicate that the difference in chemical composition between wines fermented with WE372 and that of wines fermented with VIN13, VIN13-EXS and VIN13-PPK are bigger than the differences between the two vintages. It is further interesting that once fermentations done with WE372 are removed that no pattern can be seen in the fermentations. This could be due to the relative small differences in the genetic make-up of VIN13, VIN13-EXS and VIN13-PPK and that a small change in must composition could lead to different gene expression.

3.4.1.2 Pinotage

A tabulated summary of the results from wines fermented from Pinotage grapes can be seen in Table 3.5 at the end of section 3.3.1.2.

For the FTIR data in the 2005A sample set, good discrimination is achieved. The cluster containing wines fermented with VIN13-PPK is lying a bit further away from the two clusters of wines fermented with VIN13 and VIN13-EXS.

Using FTIR data in the 2005B sample set, discrimination is good with tight clusters that are apart wide apart in the score space. The fermentations done with VIN13-EXS and VIN13-PPK seem to lie closer together which could indicate that their chemical compositions are closer to each other.

The GC data for the 2005B sample set also give good discrimination between the wines fermented with the three yeast stains. The clusters are not as tightly pack as the ones where FTIR data was used.

Using FTIR data in the 2006A sample set, good discrimination between the wines fermented with the four yeast strains are achieved. The four clusters are spread apart in score space.

Using FTIR data in the 2006B sample set we can only discriminate between wines fermented with WE372 and the rest of the wines. By further inspecting the cluster contaning wines fermented with VIN13, VIN13-EXS and VIN13-PPK we can see that there is a pattern of wines fermented with VIN13-PPK clustering together. Wines fermented with VIN13 and VIN13-EXS show some overlap. As before it can be speculated that the wines fermented with VIN13, VIN13-EXS and VIN13-PPK are more similar in chemical composition than those fermented with WE372. This could be attributed to their similarity in genotype. Using the additional set of fermentations as a test set in validation, we can see that the model is successful in discriminating the wines fermented with WE372 from the rest.

When wines fermented with WE372 is removed from sample set 2006B we can only discriminate between wines fermented with VIN13-PPK and the rest, while there seem to be no discrimination between wines fermented with VIN13 and VIN13-EXS. This would suggest that the wines fermented with VIN13 and VIN13-EXS are very similar in chemical composition. Using the test set for validation of the model does not perform very well in predicting the wines fermented with VIN13-PPK, showing some overlap between the two groups.

For the FTIR data in the 2006C sample set wines fermented with WE372 can be discriminated from wines fermented with VIN13, VIN13-EXS and VIN13-PPK. In the

group of wines that was fermented with VIN13, VIN13-EXS and VIN13-PPK, those fermented with VIN13-PPK can be separated from the rest with a small overlap with the wines fermented with VIN13 and VIN13-EXS. The clusters are not as tight as seen in sample sets 2006A and 2006B. Using the additional set of fermentations as a test set in validation, we can see that the model is successful in discriminating the wines fermented with WE372 from the rest. Discrimination can also be seen between wines fermented with VIN13-PPK and the rest although the vertical gap is not as big as with wines fermented with WE372.

Using GC data in sample set 2006C discrimination shows the same pattern as with the FTIR data. Using the GC data the clusters of wines fermented with WE372 and VIN13-PPK are lying slightly further away from the cluster containing wines fermented with VIN13 and VIN13-EXS when using FTIR data. This is confirmed in the Predicted vs. Measured plots. In the cluster containing the wines fermented with VIN13 and VIN13-EXS, no pattern can be distinguished. This could be due to a very similar chemical composition of these wines.

When combining the FTIR data from the two sample sets for 2005, no clear discrimination between the wines fermented with the three yeast strains can be seen. There is a pattern though in that the wines can be seen as three poorly defined clusters.

Even though there are no clear discrimination between the wines fermented with the four yeast strains there are three overlapping clusters which have similar patterns to those in the individual sample sets. There is a cluster with very little overlap containing wines fermented with WE372, there is a cluster with little overlap containing wines fermented with VIN13-PPK and another cluster that contains wines fermented with VIN13 and VIN13-EXS with small overlap with the rest. Using the test set for validation the model does not perform very well in predicting the wines fermented with WE372, showing overlap between the two groups.

By combining the FTIR data of all the sample sets across the two vintages we can see if it is possible to discriminate between the fermentations done with the different yeasts in successive years. This is important to create a successful model that could be used for prediction of the yeast used for a specific fermentation.

As with Clairette Blanche, the fermentations done with VIN13, VIN13-EXS and VIN13-PPK in 2005 and 2006 cluster together and could possibly be explained to be more similar in chemical composition than those fermentations done with WE372 in 2006. The fact that wines fermented with VIN13, VIN13-EXS and VIN13-PPK cluster

together for 2005 and 2006 could possibly indicate that the difference in chemical composition between wines fermented with WE372 and that of wines fermented with VIN13, VIN13-EXS and VIN13-PPK are bigger than the differences between the two vintages. As with the analysis of the individual sample sets, it appears that wines fermented with VIN13-PPK cluster together with some overlap with the wines fermented with VIN13 and VIN13-EXS. This separation of wines fermented with VIN13-PPK is much clearer using GC data.

3.4.1.3 Cabernet Sauvignon

A tabulated summary of the results from wines fermented from Pinotage grapes can be seen in Table 3.6 at the end of section 3.3.1.3.

In the 2005A sample set and using the FTIR data there is discrimination between the wines fermented with WE372 and the rest. In the other cluster there is grouping of wines fermented with VIN13 and wines fermented with VIN13-EXS and VIN13-PPK, there is some overlap between these two groups. As with most of the sample sets that include wines fermented with WE372 we can see that it forms a separate cluster away from the rest. Once the wines fermented with WE372 are removed from the sample set, the wines fermented with VIN13 form a separate cluster. Looking at the wines fermented with VIN13-EXS and VIN13-PPK there is overlap but it seems that there might be a pattern of the same wines grouping together.

Using FTIR data in the 2005B sample wines fermented with WE372 is in a separate cluster from the rest of the wines. It is not a very tight cluster and the wines are not very far from the rest of the wines. The same pattern as in the 2005A sample set can be seen in the clustering of wines fermented with VIN13, VIN13-EXS and VIN13-PPK. With the removal of the wines fermented with WE372 there is better separation between the wines fermented with the remaining three yeasts. This separation is better than in the 2005A sample set.

When using The GC data from sample set 2005B there is separation between all the wines fermented with the four different yeast strains. The wines fermented with WE372 create a very compact cluster, with the exception of two wines. It is not even necessary to remove the wines fermented with WE372 to see clustering of the wines fermented with VIN13, VIN13-EXS and VIN13-PPK.

Using FTIR data from the 2006A sample set very good discrimination was achieved between wines fermented with WE372, wines fermented with ML01 and the wines fermented with VIN13, VIN13-EXS and VIN13-PPK. The three clusters are tight and

spread far apart. The model is also very stable in prediction using the second set of fermentations as a test set in the validation process, thereby confirming the good discrimination seen in the score space. This good discrimination could possibly be explained by the differences in genotype and resulting phenotypes that result in different chemical compositions of the wines. A possible explanation for the grouping of the wines fermented with VIN13, VIN13-EXS and VIN13-PPK closely together is that their chemical profiles are very similar. This could mean that the insertion of the two foreign genes that resulted in VIN13-PPK and VIN13-EXS did not result in such a big difference in the chemical composition compared to that of the other two yeast strains.

When removing wines fermented with WE372 and ML01 from sample set 2006B wines fermented with VIN13-PPK separates from wines fermented with VIN13 and VIN13-EXS. Using the second set of fermentations as test set for validation the model does not perform well in discriminating wines fermented with VIN13-PPK from the rest as indicated by some wines having predicted values of less than 0. This could be explained by possible differences between the two sets of fermentations done for calibration of the model and the validation of the model. These differences were never analysed.

Similar results were obtained for the 2006B sample set using FTIR data as was seen in the 2005A sample set. There was good discrimination between wines fermented with WE372, wines fermented with ML01 and wines fermented with VIN13, VIN13-EXS and VIN13-PPK. Using the second set of fermentations the model performed well in discriminating the wines fermented with WE372 and ML01 from the rest of wines.

Using the GC data in the 2006B sample set the discrimination was not as good as when the FTIR data was used. Using test set validation, the model also could not discriminate between the group of wines fermented with ML01 from the others.

The same results are achieved for the 2006C sample set using FTIR data as in the 2006A and 2006B sample sets. Whereas in the 2006A and 2006B sample sets there were no discrimination possible when wines fermented with WE372 and ML01 were removed, in the 2006C sample set discrimination between wines fermented with VIN13 and wines fermented with VIN13-EXS and VIN13-PPK could be seen. The model using test set validation also result in good discrimination between the wines fermented with VIN13 an the rest of the wines. A possible explanation could be that this is due to the change in chemical composition of the wines as they age.

Using the FTIR data in the combined sample sets of 2005 no patterns could be observed in the data.

For the combination of the 2006 sample sets and using the FTIR data the same patterns as seen in the individual sample sets can be seen, three clusters representing the wines fermented with WE372, ML01 and a combination of wines fermented with VIN13, VIN13-EXS and VIN13-PPK. When using test set validation the model cannot discriminate between the groups of wines and show overlap.

Combining the 2005B and 2006B sample sets and using FTIR data, there are separate clusters for the wines fermented with WE372, wines fermented with ML01 and a combination of VIN13, VIN13-EXS and VIN13-PPK. Only wines fermented with ML01 can be discriminated from the rest of the wines when GC data are used.

3.4.2 EFFECT OF AGEING OF WINES ON DISCRIMINATION

Using GC data from the 2005B and 2006A sample sets for Clairette Blanche, two clusters can be seen representing wines samples from the two vintages.

In order to look at the effect that ageing had on the wines the score spaces that were used for discriminating the wines fermented with the different yeast strains were turned in the 3D space to reveal the structure. In most cases the resulting discrimination could possibly be explained by the chemical changes occurring during the maturation of red wine in terms of polymerisation, oxidation and other chemical reactions (Ribéreau-Gayon *et al.*, 2000).

Unlike the combined 2005 data for Cabernet Sauvignon, the combined data for 2006 using FTIR data does not show clustering for each sample set. The difference between the 2006A and the other two sample sets could very likely be attributed by the fact that the wines from 2006A sample set was sampled almost two months before the others and that it has not undergone secondary malolactic fermentation.

3.4.3 DISCRIMINATION BASED UPON NON-GM VS. GM YEAST STRAIN USED FOR FERMENTATION

With exception of this Cabernet Sauvignon 2006A sample set, all sample sets show good discrimination between wines fermented with non-GM and GM yeast when looking at score plots for each PLS1-discrimination model (results not shown). Discrimination between wines fermented with non-Gm and GM yeast is not good when looking at the combined sample sets for 2005 and for the samples sets of 2006 when using FTIR data. More data is needed to further look at the possibility of using FTIR data for discrimination combining vintages.

3.5 CONCLUSIONS

Regarding the first objective of this study, it can be stated that the possibility exist for using FTIR and GC for the discrimination of wines fermented with different yeast strains within each cultivar. Generally the discrimination for each sample set in a specific year shows the same discrimination in the alternative year. When combining the sample sets for a specific year, discrimination is still possible between wines fermented from non-similar yeast. In most cases it is also possible to discriminate between the different sample sets taken at different times within a vintage.

The second objective of the study resulted in discrimination of wines fermented with different yeast strains by using FTIR or GC data. The results showed that FTIR is outperforming GC with certain sample sets and *vice versa* for other sample sets. Overall the discrimination with FTIR or GC data was the same. Discrimination was also tested by using PLS2-discrimination with FTIR and GC data combined. This combination of data did not significantly increase discrimination between wines fermented with different yeasts. With the time saving benefits of FTIR it would perhaps be beneficial investigating this as technique of preference in further studies.

It is clear that more data is needed to draw further conclusions regarding the discrimination of wines fermented with closely related yeast strains (VIN13, VIN13-EXS and VIN13-PPK).

In none of the chemometric analysis of the wines did the addition of a commercial enzyme show to have any difference in the discrimination ability. It can therefore be speculated that the influence of enzyme addition is limited.

When looking at the PLS1-discrim results for discrimination of the GM and non-GM fermented wines, the FTIR technique could lead to a possible fast authentication technique when screening wines for the European Union market to ensure no GM material is used. In order for this possibilty to be developed into an accepted method however, much more work would be needed to establish reliable, robust models. A comprehensive industry-wide calibration must include all major S.A. cultivars as well as further variance-inducing factors, such as maturation of wine, origin of grapes and other factors that could affect the wine. This will no doubt be a considerable project, but the entire foundation has been worked out for the first time in this study.

3.6 FUTURE STUDIES

The present feasibility study prediction models can easily be extended to any number of additional cultivars needed or wished for. To further improve the robustness, wines made from employing different process technologies, blending of cultivars and mixtures of yeasts should also be included. The study should also be extended to include the use of Near Infrared technology which is widely used in the food and beverage industry for very fast, on-line process analytical chemistry (PAC) applications.

A related advantage is that the S.A. wine industry will be able to monitor processes based on *at-line* NIR or FTIR Process Analytical Technologies (PAT) (McLennan and Kowalski, 1995, Bakeev, 2005). There are also numerous promising potentials and perspectives for a new generation approach to *on-line* process sampling in the industrial winemaking arena (Esbensen *et al.*, 2006), which is intimately connected to PAT.

3.7 LITERATURE CITED

Anonymous, 2003. Regulation (EC) No 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC. Official Journal of the European Union 46 24-28

Arvanitoyannis, I.S., Katsota, M.N., Psarra, E.P., Soufleros, E.H., Kallithraka, S., 1999. Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). Trends Food Sci Technol 10 (10), 321-336

Bakeev, K., 2005 Process Analitycal Technology Spectroscopic Tools & Implementation. Blackwell Publishing, Oxford

Boulton, R.B., Singleton, V.L., Bisson, L.F., Kunkee, R.E., 1995 (1st ed). Principles and practices of winemaking. Chapman Hall, New York

Cerna, M., Barros, A.S., Nunes, A., Rocha, S.M., Delgadillo, I., Copikova, J., Coimbra, M.A., 2003. Use of FT-IR spectroscopy as a tool for the analysis of polysaccharide food additives. Carbohydr Polym 51 (4), 383-389

Cozzolino, D., Smyth, H.E., Gishen, M., 2003. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. J Agric Food Chem 51 (26), 7703-7708

Duarte, I.F., Barros, A., Almeida, C., Spraul, M., Gil, A.M., 2004. Multivariate analysis of NMR and FTIR data as a potential tool for the quality control of beer. J Agric Food Chem 52 (5), 1031-1038

Edelmann, A., Diewok, J., Schuster, K.C., Lendl, B., 2001. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. J Agric Food Chem 49 (3), 1139-1145

Engelhard, S., Löhmannsröben, H.-G., Schael, F., 2004. Quantifying ethanol content of beer using interpretive near-infrared spectroscopy. Appl Spectrosc 58 (10), 1205-1209

Esbensen, K.H., 2002 (5th ed). Multivariate Data Analysis - in Practice CAMO Process AS, Oslo

Esbensen, K.H., Friis-Petersen, H.H., Petersen, L., Holm-Nielsen, J.B., Mortensen, P.P., 2006. Representative process sampling – in practise: Variographic analysis and estimation of Total Sampling Errors (TSE). In: Chemometrics and Intelligent Laboratory Systems. Proceedings 5th Winter Symposium of Chemometrics (WSC-5), 2006, Samara, Russia, (accepted for publication)

European Commission, 2003a. Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. OJ Eur Comm 46 1-23

European Commission, 2003b. Regulation (EC) No 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC. OJ Eur Comm 46 24-28

Faria, M.A., Magalhaes, R., Ferreira, M.A., Meredith, C.P., Monteiro, F.F., 2000. *Vitis vinifera* must varietal authentication using microsatellite DNA analysis (SSR). J Agric Food Chem 48 (4), 1096-1100

Garcia-Beneytez, E., Moreno-Arribas, M.V., Borrego, J., Polo, M.C., Ibanez, J., 2002. Application of a DNA analysis method for the cultivar identification of grape musts and experimental and commercial wines of Vitis vinifera L. using microsatellite markers. J Agric Food Chem 50 (21), 6090-6096

Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R., 1992. Simultaneous amplification and detection of specific DNA sequences. Nat Biotech 10 (4), 413-417

Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. Nat Biotech 11 (9), 1026-1030

Klein, D., 2002. Quantification using real-time PCR technology: applications and limitations. Trends Mol Med 8 (6), 257-260

Lambrechts, M.G., Pretorius, I.S., 2000. Yeasts and its importance to wine aroma. S Afr J Enol Vitic 21 97– 129

Leopold, S., Uehlein, N., Kaldenhoff, R., Schartl, A., 2003. Fate of DNA during must fermentation. Acta Hort 603 133-134

Martens, H., Naes, T., 1998 Multivariate Calibration. Wiley and Sons Ltd., Chichester

McLennan, F., Kowalski, B.R., 1995 Process analytical chemistry. Chapman & Hall, London

Palma, M., Barroso, C.G., 2002. Application of FT-IR spectroscopy to the characterisation and classification of wines, brandies and other distilled drinks. Talanta 58 (2), 265-271

Patz, C.D., Blieke, A., Ristow, R., Dietrich, H., 2004. Application of FT-MIR spectrometry in wine analysis. Anal Chim Ac 513 (1), 81-89

Petersen, L., Esbensen, K.H., 2005. Sampling in Practice - A TOS Toolbox of Unit Operations. In: Pomerantsev, A. L. (ed). Progress in chemometric research . Nova publishers, New York. pp. 245-263

Petersen, L., Minkkinen, P., Esbensen, K.H., 2005. Representative sampling for reliable data analysis: Theory of Sampling. Chemom Intell Lab Syst 77 (1-2), 261-277

Pretorius, I.S., Bauer, F.F., 2002. Meeting the consumer challenge through genetically customized wine-yeast strains. Trends Biotechnol 20 (10), 426-432

Reid, L.M., O'Donnell, C.P., Downey, G., 2006. Recent technological advances for the determination of food authenticity. Trends Food Sci Technol 17 (7), 344-353

Ribéreau-Gayon, J., Glories, Y., Maujean, A., Dubourdieu, D., 2000. The chemistry of wine stabilization and treatments. In: Ribéreau-Gayon, J. (ed). Handbook of Enology, 2. Wiley, Chichester, England. pp. 404

Romano, P., Fiore, C., Paraggio, M., Caruso, M., Capece, A., 2003a. Function of yeast species and strains in wine flavour. Int J Food Microbiol 86 (1-2), 169-180

Romano, P., Granchi, L., Caruso, M., Borra, G., Palla, G., Fiore, C., Ganucci, D., Caligiani, A., Brandolini, V., 2003b. The species-specific ratios of 2,3-butanediol and acetoin isomers as a tool to evaluate wine yeast performance. Int J Food Microbiol 86 (1-2), 163-168

Romano, P., Suzzi, G., Domizio, P., Fatichenti, F., 1997. Secondary products formation as a tool for discriminating non-*Saccharomyces* wine strains. A Van Leeuw 71 (3), 239-242

Savazzini, F., Martinelli, L., 2006. DNA analysis in wines: Development of methods for enhanced extraction and real-time polymerase chain reaction quantification. Anal Chim Ac 563 (1-2), 274-282

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least-squares procedures. Anal Chem 36 (8), 1627-1639

Scotter, C.N.G., 1997. Non-destructive spectroscopic techniques for the measurement of food quality. Trends Food Sci Technol 8 (9), 285-292

Siret, R., Boursiquot, J.M., Merle, M.H., Cabanis, J.C., This, P., 2000. Toward the authentication of varietal wines by the analysis of grape (*Vitis vinifera* L.) residual DNA in must and wine using microsatellite markers. J Agric Food Chem 48 (10), 5035-5040

Siret, R., Gigaud, O., Rosec, J.P., This, P., 2002. Analysis of grape Vitis vinifera L. DNA in must mixtures and experimental mixed wines using microsatellite markers. J Agric Food Chem 50 (13), 3822-3827

WineScan FT120 Type 77110 and 77310 Reference Manual, Issue 4 GB Foss Electric, Denmark, 2001, http://www.foss.dk

Strauss, M., 2003, The heterologous expression of polysaccharase genes with enological relevance in Saccharomyces cerevisiae. . MSc Thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa

Tzouros, N.E., Arvanitoyannis, I.S., 2001. Agricultural Produces: Synopsis of Employed Quality Control Methods for the Authentication of Foods and Application of Chemometrics for the Classification of Foods According to Their Variety or Geographical Origin Crit Rev Food Sci Nutr 41 (4), 287- 319

Wilson, R.H., Colquhoun, I.J., Kemsley, E.K., 2001. Screening food products for authenticity and adulteration: the use of infrared spectroscopy. Am Lab 33 (11), 54-55

Chapter 4

RESEARCH RESULTS

The use of Fourier transform infrared (FTIR) spectroscopy for yeast strain phenotype discrimination

The use of Fourier transform infrared (FTIR) spectroscopy for yeast strain phenotype discrimination

C.D. Osborne[1], K.H. Esbensen[1,2], P van Rensburg[1], H.H. Nieuwoudt[1]

1. IWBT (Institute for Wine Biotechnology), Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa

2. ACABS (Applied Chemometrics, Analytical Chemistry, Bioenergy, Applied Biotechnology & Sampling) research group, Aalborg University Esbjerg (AAUE), Denmark

Abstract

In this study FTIR was used in transmission and in attenuated total reflectance (ATR) modes to discriminate between phenotypes of *Saccharomyces cerevisiae* yeast. Both FTIR modes resulted in positive discrimination between phenotypes and between closely related genetically modified yeast. The FTIR-ATR was conducted by using a Type IIa diamond crystal. Further work is required in sample preparation and more yeast strains should be included in the FTIR-transmission studies to validate the PLS-discrimination models.

Keywords: Mid-infrared spectroscopy; FTIR-ATR, FTIR-transmission, Chemometrics; PLS-discrim; *Saccharomyces cerevisiae* yeast

4.1 INTRODUCTION

Fourier transform infrared (FTIR) spectroscopy has been used for the identification and discrimination of bacteria as far back as the 1950's and 1960's, but was not pursued as a viable method due to the limitations of available technology at the time (for an in depth review see Naumann *et al.*, 1991b). With the advancement of infrared instrumentation, more powerful computers and advanced algorithms for multivariate data analysis and pattern recognition, FTIR as a tool has become more widely acceptable and used (Mariey *et al.*, 2001).

The use of FTIR spectroscopy in biological systems is based on the amount of infrared energy absorbed by the infrared active molecules in a sample. The absorption leads to increased vibrations in the highly polar bonds of functional groups. The resulting absorption, transmission or reflection spectrum gives information about the total biochemical composition of a sample regarding the molecular composition (Zhao *et al.*, 2006). FTIR can be seen as a rapid, whole organism fingerprint approach (Zhao *et al.*, 2006; Naumann *et al.*, 1991b). These spectral fingerprints can be highly specific and can be used as a measure of the phenotype of an organism in taxonomic classification (Timmins *et al.*, 1998b), down to the strain and serogroup/serotype level (Ngo Thi *et al.*, 2000).

The spectral fingerprint of the whole cell is made up of several components. There are two major components of the cell, namely the cell membranes (cell wall and cytoplastic membrane) which mainly consist of lipids and polysaccharides and the cytoplast which mainly consist of nucleic acids and proteins (Naumann, 1998; Yu and Irudayaraj, 2005). There is some general agreement on the spectral absorption bands for identification of the different components (Maquelin *et al.*, 2002; Yu and Irudayaraj, 2005), but there is still uncertainty of the exact regions for the different components. Table 4.1 list spectral absorption for components in the infrared region (Maquelin *et al.*, 2002). For protein structure analysis the amide-I band, arising from the peptide backbone, is mostly used (Wolkers and Oldenhof, 2005). The FTIR spectra of bacteria are mostly dominated by the amide-I and amide-II bands of the various proteins and oligo- and polysaccharides of the cell wall (Naumann *et al.*, 1995).

For identification and characterisation of microorganisms by FTIR it is not necessary to identify the bands and their intensities of the different components in the spectra. The whole fingerprint can be used in conjunction with chemometrics for identification purposes (Maquelin *et al.*, 2002). Data analysis can be made more effective by using only the 1440-720cm^{-1} area of the spectra as this covers the mixed region of proteins

and fatty acids, the polysaccharides present in the cell wall, and the “true” chemical fingerprint region where the various bands can not be assigned to specific functional groups (Oust *et al.*, 2004a).

Table 4.1: Assignment of absorbance bands in the IR spectra of microbial cells (Maquelin *et al.*, 2002)

Frequency (cm^{-1})	Assignment[a]
3500	O–H str of hydroxyl groups
3200	N–H str (amide A) of proteins
2955	C–H str (asym) of $–CH_3$ in fatty acids
2930	C–H str (asym) of $>CH_2$
2918	C–H str (asym) of $>CH_2$ in fatty acids
2898	C–H str of C–H in methine groups
2870	C–H str (sym) of $–CH_3$
2850	C–H str (sym) of $>CH_2$ in fatty acids
1740	>C=O str of esters
1715	>C=O str of carbonic acid
1680– 1715	>C=O in nucleic acids
1695, 1685, 1675	amide I band components resulting from antiparallel pleated sheets and β-turns of proteins
1655	amide I of α-helical structures
1637	amide I of β-pleated sheet structures
1550– 1520	amide II 1515 ‘‘tyrosine’’ band
1468	C–H def of $>CH_2$
1400	C=O str (sym) of COO^-
1310– 1240	amide III band components of proteins
1250– 1220	P=O str (asym) of $>PO_2^-$ phosphodiesters
1200– 900	C– O, C–C str, C–O–H, C–O–C def of carbohydrates
1090– 1085	P=O str (sym) of $>PO_2^-$
720	C–H rocking of $>CH_2$
900– 600	“fingerprint region”

[a] str=stretching; def=deformation; sym=symmetric; asym=asymmetric.

To ensure the generation of good quality data it is very important that FTIR spectroscopy measurements are reproducible. There are several factors that can influence the reproducibility including cell cycle, growth stage of the cells, growth conditions, sampling and sample preparation. It is vital that all these parameters are controlled (Maquelin *et al.*, 2002). It has been shown that FTIR spectra are robust against small changes in bacterial growth conditions and that larger variations in the growth medium influenced the separation of strains (Oust *et al.*, 2004a). The physical state of a sample, for example during the preparation of KBr pellets (such as hydration, pressure it was exposed to, grinding and mixing) can have a severe influence on FTIR results (Wolkers and Oldenhof, 2005; Naumann, 1998). A standard protocol for sample

preparation that is used by several groups has been published (Maquelin *et al.*, 2002) after it was first proposed by Naumann *et al.*, 1991a).

FTIR has been used to differentiate, classify and identify *Lactobacilli* (Oust *et al.*, 2004a; Oust *et al.*, 2004b), *Listeria* strains (Lin *et al.*, 2004), sulphate and thiosulphate reducing bacteria (Rubio *et al.*, 2006), *Candida* (Essendoubi *et al.*, 2005; Tintelnot *et al.*, 2000; Timmins *et al.*, 1998a), *Saccharomyces cerevisiae* (Wenning *et al.*, 2002; Timmins *et al.*, 1998b) and *Streptomyces spp.* (Zhao *et al.*, 2006).

The aim of the study was to evaluate the effectiveness of FTIR in transmission mode and FTIR in ATR (attenuated total reflectance) mode to discriminate between different *Saccharomyces cerevisiae* strains. For the FTIR in transmission mode, the discrimination of yeast strains was investigated where yeast samples were suspended in a yeast-peptone-dextrose (YPD) medium and in a water medium. For the FTIR in ATR mode, the discrimination of yeast strains were investigated with active dried wine yeast (ADWY) in a powder from and in a pellet form and with powdered dried yeast made from liquid cultured yeast.

4.2 MATERIALS AND METHODS

4.2.1 INSTRUMENTATION

4.2.1.1 FTIR - transmission analysis

Infrared analysis was carried out in the mid-infrared region (MIR)(5011 to 929 cm^{-1} at 4 cm^{-1} intervals) with a WineScan FT120 instrument (Foss Electric, Denmark). Samples are pumped through a CaF_2-lined cuvette with an optical path length of 37 µm. Samples are preheated to 40℃ in a heater block before analysis. Each sample was analysed in duplicate, each spectrum was composed by the average of twenty scans after being Fourier transformed. Foss Zero Liquid S-6060 (WineScan FT120 Reference Manual) was analysed before samples to facilitate correction for the specific background spectrum (especially water) present. The final spectra were generated by rationing the individual sample spectrum by the zero solution spectra at each recorded wavelength. FTIR data were recorded as absorbencies as a function of wavenumber (cm^{-1}).

4.2.1.2 FTIR - ATR analysis

Infrared analysis was carried out in the mid-infrared region (MIR)(4000 to 650 cm^{-1} at 4 cm^{-1} resolution) with a NEXUS 670 FT-IR (Thermo Electron, USA) instrument. The spectrometer equipped with a Ge-on-KBr beamsplitter and DTGS/CsI detector, was continuously purged with UHP nitrogen gas (AFROX, South Africa). The spectrometer

was fitted with a Smart Golden Gate Single Reflection Diamond ATR sampling accessory featuring a type IIa diamond. Using the Smart Golden Gate accessory, samples require little or no preparation with the only requirement being that the sample must be in optimal contact with the diamond. Uniform contact is achieved by using a self-levelling anvil with a sapphire tip that can exert pressures up to 17.6 kg/cm^2. Properties of the type IIa diamond are listed in Table 4.2 (Anonymous, 2002). Fig. 4.1 shows the %transmittance spectrum for the Type IIa diamond with the spectral bands at 2600 to 1600 cm^{-1}. Each sample spectrum is composed by the average of twenty scans after being Fourier transformed. No ATR path length correction was done as spectra were not used for quantification. The operating and data manipulation was performed with the basic OMNIC (Thermo Electron Inc., San Jose, CA, USA) software package for spectroscopy. Residual water or CO_2 absorptions were removed by spectral subtraction with the OMNIC spectroscopy software. Background scans were performed after every fifth sample scan.

Table 4.2: Type IIa diamond specification (Anonymous, 2002)

Transmission Range	4500 to 2500 cm^{-1} & 1667 to 33 cm^{-1}
Refractive Index	2.4
% Transmittance (thickness)	70% @ 1 mm thickness
Cleaning Agents	alcohol, acetone, H_2O
Solubility In Water (100g H_2O @25℃)	insoluble
Density g/cm^3	3.51
Solvents Which Attack	$K_2Cr_2O_7$, conc H_2SO_4
Max Temp In Air [℃]	750
Melting Point [℃]	3500
Hardness [kg/mm^2] (Knoop #):	7000
Composition	crystalline carbon, single crystal
Crystal Class	cubic
pH Range	1-14
Depth of Penetration*	2.01

* Assumes a 45 degree crystal and 1000 cm^{-1} with a sample refractive index of 1.5

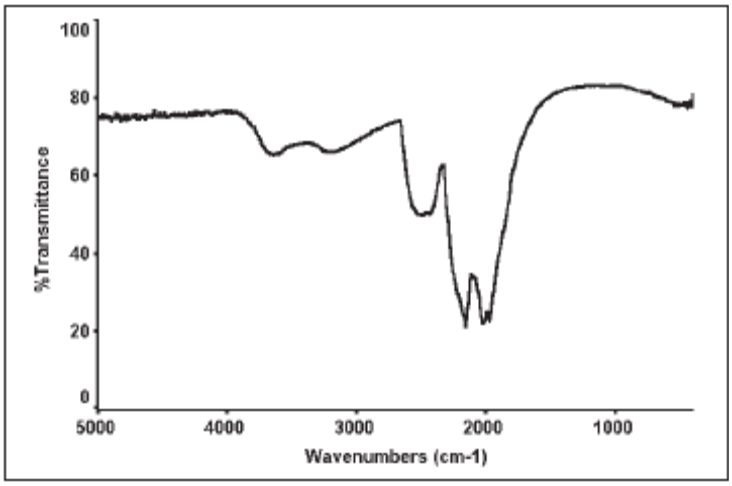


Figure 4.1: %Transmittance spectra for Type IIa diamond (Anonymous, 2002)

4.2.2 YEAST STRAINS AND GROWTH CONDITIONS

In order to test the differences between YPD autoclaved in different autoclaves, YPD broth was made-up and autoclaved in two different autoclaves at 120℃ for 15min. Yeast was cultivated, as described earlier, and was diluted with YPD to achieve a final OD_{600} of 0.6 and 1.1 respectively.

S. cerevisiae yeast strains were taken from the freeze culture collection of the Institute of Wine Biotechnology, Stellenbosch University. For the use with FTIR-transmission, VIN13 and WE372 yeast strains and for the use with FTIR-ATR, VIN13, VIN13-EXS, VIN13-PPK, VIN13-DLG29 and VIN13-DLG30 were used for the experiments (Table 2). For the experiments using FTIR-ATR, different production batches were also used, see Table 4.3 for the production batch numbers.

Strains were streaked out on YPD agar plates (1% yeast extract, 2% peptone, 2% dextrose, 2% agar, Biorad, South Africa) from freeze cultures and incubated at 30℃ for 48 hours. A single colony was picked and inoculated in a 10 mL YPD liquid (1% yeast extract, 2% peptone and 2% dextrose, Biorad, South Africa) culture. Liquid cultures were grown for 48 hours at 30℃ on a rotating wheel. At least five independent repeats of each strain were grown for each experiment.

Table 4.3: Yeast strains used in this study

***S. cerevisiae* strain**	**Batch number**	**Type[a]**	**Genotype/Description**	**Source/Reference**
VIN13	n/a	Freeze	Commercial diploid strain	Anchor Yeast Technologies (SA)
VIN13-EXS	n/a	Freeze	*ura3::ADH1$_P$-MFα1$_S$-END1-TRP5$_T$-ADH1$_P$-XYNC-ADH2$_T$*	Strauss, 2003
VIN13-PPK	n/a	Freeze	*ura3::ADH1$_P$-MFα1$_S$-PELE-TRP5$_T$-ADH1$_P$- MFα1$_S$-PEH1-TRP5$_T$*	Strauss, 2003
VIN13-DLG29	n/a	Freeze	*ILV2::TEF1$_P$-XYN2-ADH2$_T$*	Louw, 2004
VIN13-DLG30	n/a	Freeze	*ILV2::ADH1$_P$-MFα1$_S$-END1-TRP5$_T$*	Louw, 2004
WE372	n/a	Freeze	Commercial strain	Anchor Yeast Bio-Technologies (SA)
VIN13	V4-26J	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
VIN13	V5-59D	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
WE372	V4-41	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
WE372	V5-46D	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
WE372	V5-47D	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
NT7	V1-30	ADWY	Commercial strain	Anchor Yeast Bio-Technologies (SA)
AWRI R2	412113	ADWY	Commercial strain	AB Mauri (Australia)
Maurivin B	516106	ADWY	Commercial strain	AB Mauri (Australia)

a: Freeze – Freeze culture, 60% Glycerol; ADWY – Active Dried Wine Yeast

4.2.3 SAMPLE PREPARATION

4.2.3.1 Samples for FTIR transmission

Samples were presented to the FTIR in two liquid mediums, YPD and H_2O respectively.

Preparations of yeast suspended in YPD were made. Optical density at 600nm (OD_{600}) was measured (Helios Beta, Thermo Electron Corp, England) for the yeast cultured in YPD. Samples were diluted into 25 mL YPD to achieve a predetermined (05 and 1.0) OD_{600} value. Using Formula 4.1, the volume of cultured yeast suspension needed to be diluted in YPD to achieve the desired OD600 was calculated:

$$V_i = V_f \bullet \frac{OD_i}{OD_d} \qquad (4.1)$$

Where

V_i = volume needed of cultured yeast suspension

V_f = final volume of yeast suspension at desired OD_{600}

OD_i = initial measured OD_{600} of cultured yeast suspension

OD_d = desired OD_{600} of the final yeast suspension

The diluted 25 mL samples of yeast suspended in YPD were used for FTIR analysis.

Preparations were made of yeast suspended in de-ionised water. The 10 mL liquid cultures were concentrated by centrifugation at 2320xg for 30 seconds (Eppendorf 5415D, Hamburg Germany). The pellets were then resuspended and washed three times in sterile de-ionised water. The washed pellet was resuspended in 1 mL sterile de-ionised water. The resuspended yeast was then added to 25 mL sterile de-ionised water. After the dilution the OD_{600} was determined (Helios Beta, Thermo Electron Corp, England). The way the dilution was done resulted in a wide range of OD_{600} values; the statistics is shown in Table 4.4.

Table 4.4: Statistics for Low and High OD yeast/water data sets

Yeast	Average OD600	Min	Max	Std	n
VIN13	2.09	0.93	2.63	0.52	41
WE372	1.98	0.86	2.55	0.56	29

Min: minimum OD600; Max: maximum OD600; Std: standard deviation; n: number of samples

4.2.3.2 Samples for FTIR-ATR

Samples were prepared from Active Dried Wine Yeast (ADWY) and from liquid cultures. Samples were presented in two forms, as hydraulically pressed pellets and as fine powders.

4.2.3.2.1 Active Dried Wine Yeast (ADWY)

ADWY is packaged in granular form in vacuum sealed packs. Granules were crushed in an onyx mortar and pestle and then ground down to a fine powder, this sample preparation was repeated independently on two different days. The powder was presented to the ATR crystal of the FTIR instrument. Powder was also pressed into pellets. A powder sample of roughly 0.1g was weighed off and transferred to a stainless steel holder (13 mm inside diameter). The holder was placed in a hydraulic press and evacuated with an attached vacuum pump. A pressure of about 10000 kg (roughly 7500 kg/cm^2) was applied for two minutes. The pellet was presented to the ATR crystal. To ensure that cracking of the pellet was minimised a Perspex holder was placed around the pellet.

4.2.3.2.2 Yeast from liquid cultures

For these experiments five yeast strains were used, namely VIN13 and four genetically modified stains of VIN13. Two of these strains (VIN13-PPK and VIN13-EXS) have each two foreign genes inserted into their genome while the other two strains (VIN13-DLG29 and VIN13-DLG30) each only have one foreign gene inserted into their genome. The

potential of PLS-discrimination will be tested by the selected five strains as they closely related in terms of genetic make-up.

The 10 ml liquid yeast cultures, cultured as described earlier, were concentrated by centrifugation at 2320xg for 30 sec (Eppendorf 5415D, Hamburg Germany). The pellets were resuspended in sterile de-ionised water after which it was washed three times in sterile de-ionised water. The washed pellet was resuspended in 1 mL sterile de-ionised water and transferred to a Petri dish. The yeast suspension was dried overnight at 55℃. The resulting dried yeast flake was ground to a fine powder in an onyx mortar and pestle. The powder was presented to the ATR crystal of the FTIR instrument.

4.2.4 CHEMOMETRIC DATA ANALYSIS

Corrected (time and date stamps and other non-spectral data were removed from the raw data) FTIR spectra were imported in the Unscrambler software (version 9.2, Camo PROCESS AS, Oslo, Norway) for PLS-discrimination.

Pre-processing was carried out on all spectra. The second derivative was employed on FTIR transmission and FTIR ATR spectra using the Savitzky-Golay algorithm (Savitzky and Golay, 1964) with five data points left – and right, with a second-order polynomial fit. The second order derivatives were used to compensate for baseline shifts, to remove possible minor scattering effects and for peak sharpening to bring out small peaks and changes in slope. Finally the data were mean centred and standardised by dividing each column in the data matrix with the standard deviation of each variable (wavenumber). This combined treatment is termed auto-scaling in chemometrics (Esbensen, 2002).

Data analysis on spectra from the FTIR-ATR work was done on a reduced spectral region (1440 to 720 cm^{-1}) after an initial survey of the data of the full spectral range (4000 to 650 cm^{-1}). The reduced spectral region covers the mixed region of proteins and fatty acids, the region covering polysaccharides present in the cell wall, and the "true" chemical fingerprint region where the various bands can not be assigned to specific functional groups (Oust *et al.*, 2004a).

PLS2-discriminant analysis was used for discriminating between the different yeast strains. A dummy Y variable was created to represent each of the yeast strains used. A value of +1 was assigned to a sample when it belonged to a specific yeast class and -1 if it did not belong to that class. These PLS2-discrimination models were used to analyse the data structure and not for prediction; due to this leverage correction was used as a validation method.

PLS1-discriminant analysis was used with data generated with FTIR-transmission where only two yeast strains were used, VIN13 and WE372. A dummy Y variable was created representing either WE372 (signified by: -1) and VIN13 (signified by: +1).

Full description of PLS-discrimination can be found in standard chemometric textbooks (Esbensen, 2002, Martens and Naes, 1998).

4.3 RESULTS AND DISCUSSION

4.3.1 DISCRIMINATION OF YEAST STRAINS BY USING FTIR-TRANSMISSION

4.3.1.1 Suspended in YPD

Fig. 4.2a shows the score plot for the PLS2-discrim done with samples of YPD broth and cultivated yeast at the two indicated OD_{600} values. Leverage correction was used as validation method. The samples of YPD broth form one cluster which indicates that the different treatments in different autoclaves did not have a significant effect on the clustering of these samples. It is therefore possible that preparation of YPD is not of major importance as long as the same procedure for make up is followed. The first PLS component (PC1) describes the change in optical density of the samples, and those with a low OD located to the right and those samples with higher OD to the left of PC1. The second PLS component (PC2) describes the yeast strain, and VIN13 located to the positive end and WE372 located to the negative end, YPD broth, which is the diluting medium for the two yeast strains, located on the zero line between the groups of the two yeast strains.

Fig. 4.2b represents a three dimensional view of the first three PLS components (PC1, PC2, PC3) in the score space shown in Fig. 2, but rotated to reveal new features in the data. Three clusters can be identified representing samples of YPD broth, VIN13 samples and WE372 samples, respectively.

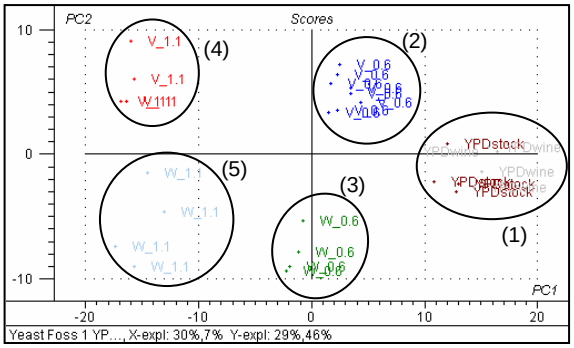


Figure 4.2a: PLS2-Discrim score plot for yeast scanned in FTIR-transmission in YPD as liquid medium. Objects represent YPD (●,●) with no yeast (1); VIN13 (●) (2) and WE372 (●) (3) at OD_{600} 0.6; VIN13 (●) (4) and WE372 at OD_{600} 1.1 (●) (5).

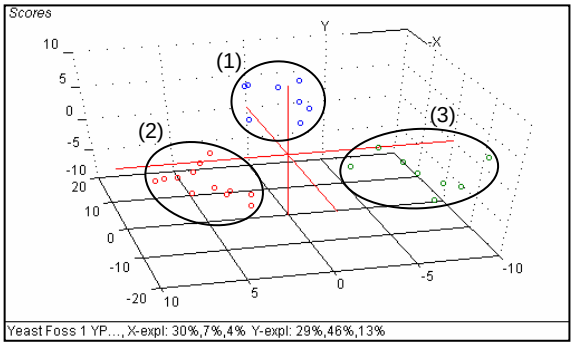


Figure 4.2b: PLS2-Discrim score plot for yeast scanned in FTIR-transmission in YPD as liquid medium. Samples have OD_{600} of 0.6 and 1.1. **o** – YPD (1); **o** - VIN13 (2); **o** - WE372 (3)

If future prediction models must be created, a PLS1-discrimination model for each yeast strain must be created, while test set validation must be used.

4.3.1.2 Suspended in water

Data of generated spectra for five independent sets of experiments, done on different days were combined for the data analysis.

Fig. 4.3 shows the score plot for the PLS1-discrim with leverage correction as a method of validation. There are two clusters representing the diluted samples of VIN13 and the diluted samples of WE372. It appears that the optical density of the samples does not play an important part in the discrimination of the samples of the two yeast strains, as both groups of yeast contain a range of samples with different optical densities.

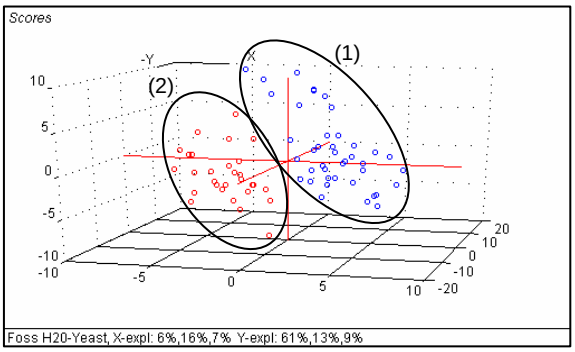


Figure 4.3: PLS1-discrim score plot for yeast scanned in FTIR-transmission in H_2O as liquid medium. Leverage correction used. OD values between 0.86-2.63.
o VIN13 (1); **o** WE372 (2)

4.3.2 DISCRIMINATION OF YEAST USING FTIR-ATR

4.3.2.1 Active dried wine yeast (ADWY)

Fig. 4.4a shows the score plot of the PLS2-discrim for all ADWY samples (see Table 2). Leverage correction was used as validation method. There are seven clusters of samples representing Maurivin B, AWRI R2, NT7, VIN13 produced in 2004, VIN13 produced in 2005, WE372 produced in 2004 and the combined two batches of WE372 produced in 2005, respectively. For both VIN13 and WE372 there are separate clusters for the different production years. A possible explanation for this could be that there were slight differences and variations in the production processes that could result in different phenotypes of the different yeasts. If future prediction models must be created, a PLS1-discrimination model for each yeast strain must be created, while test set validation must be used.

Fig. 4.4b shows the score plot of the discrimination of the yeast strains using the yeast strains as Y-variables, it is the same score plot as presented in Fig. 4.4a, the samples are identified by the way they were presented to the ATR crystal (either granular, pellet or powder). It can be observed that within the clusters there is no pattern of how the samples were presented to the ATR crystal. The different forms of sample presentation did not have an effect on the discrimination of the different yeasts,

with the exception of granular samples in the WE372, Bx 1 (2004) cluster, which shows some separation from the powdered samples in the cluster.

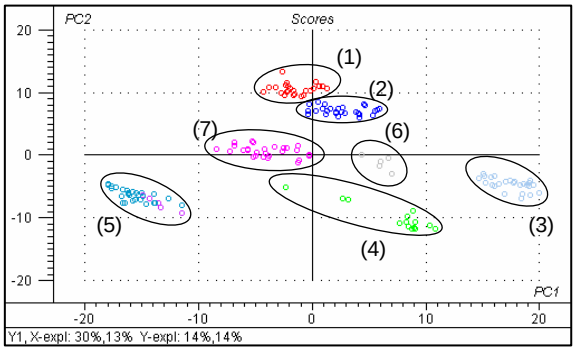


Figure 4.4a: PLS2-Discrim score plot for ADWY yeast scanned in FTIR-ATR with different presentation forms. Leverage correction used. o – Maurivin B (1); o – AWRI R2 (2); o – NT7 (3); o – WE372 (Bx 1, 2004) (4); o – WE372 (Bx 2, 2005) (5); o – WE372 (Bx 3) (5); o – VIN13 (Bx 1, 2004) (6); o – VIN13 (Bx 2, 2005) (7)

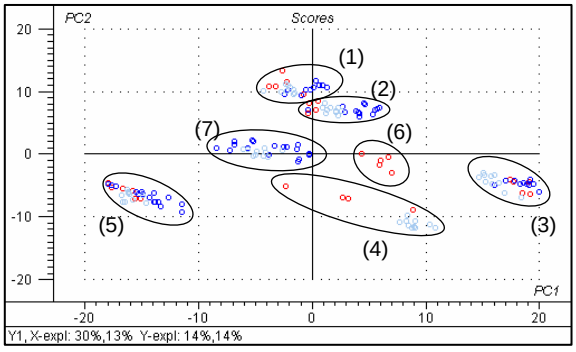


Figure 4.4b: PLS2-Discrim score plot for ADWY yeast scanned in FTIR-ATR with different presentation forms. Leverage correction used. o – granuals; o – pellet; o – finely ground powder (1) Maurivin B; (2) AWRI R2; (3) NT7; (4) WE372 (Bx 1, 2004); (5) WE372 (Bx 2, 2005); (5) WE372 (Bx 3); (6) VIN13 (Bx 1, 2004); (7) VIN13 (Bx 2, 2005)

4.3.2.2 Yeast from liquid cultures

Fig. 4.5 shows the score plot of five yeast strains. Leverage correction was used as a method of validation. There are five clusters, each representing each of the five yeast strains used. Discrimination of the five yeast strains are possible when using FTIR-ATR as analytical instrument and using yeast in a powder form as sample presentation to the ATR crystal.

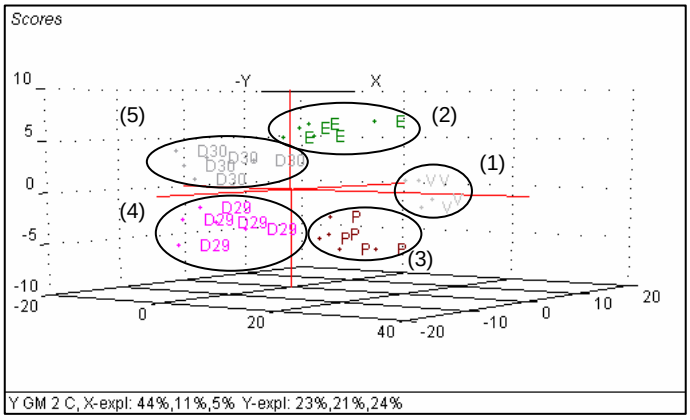


Figure 4.5: PLS2-Discrim score plot for yeast scanned in FTIR-ATR as a powder. Leverage correction used. ● – VIN13 (1); ● – VIN13-EXS (2); ● – VIN13-PPK (3); ● – VIN13-DLG29 (4); ● – VIN13-DLG30 (5)

4.4 CONCLUSIONS

It was shown in this study that it is possible to use both FTIR-transmission and FTIR-ATR to discriminate between different yeast strain phenotypes. It was shown that when using FTIR-transmission there is discrimination between yeast samples when the yeast was suspended in YPD and in water. Dried yeast samples could be discriminated when the yeast were in a granular, powder form or in a pellet form, using FTIR-ATR. It was

possible with FTIR-ATR to discriminate between the closely related yeast strain phenotypes.

FTIR-transmission and FTIR-ATR are both fast techniques and these methods could be employed in fast screening of yeast phenotypes in a laboratory environment. With slight modifications to instrumentation to fit into a production area, these methods could be applied at-line or even in-line in a yeast production process to ensure consistency in product purity and quality through the application PAT (Process Applied Technology).

For the general discrimination of yeast based on their chemical fingerprint, the type IIa is a suitable crystal to use in FTIR-ATR. Comparative analysis with other available ATR crystals (germanium, zinc selenide, etc.) should also be investigated for suitability as some of them have lower absorption in the amide I and II regions (Anonymous, 2002) and are generally cheaper.

Future work should also include FTIR-ATR work that compares yeast collected off agar plates in a dry and wet state. Further work is also needed in the standardisation of the methods and quantification of the robustness of sample preparation.

4.5 LITERATURE CITED

Anonymous, 2002 Crystal reference guide. Thermo Electron Spectroscopy, Madison, WI

Esbensen, K.H., 2002 (5th ed). Multivariate Data Analysis - in Practice CAMO Process AS, Oslo

Essendoubi, M., Toubas, D., Bouzaggou, M., Pinon, J.-M., Manfait, M., Sockalingum, G.D., 2005. Rapid identification of *Candida* species by FT-IR microspectroscopy. Biochim Biophys Acta 1724 (3), 239-247

Lin, M., Al-Holy, M., Al-Qadiri, H., Kang, D.H., Cavinato, A.G., Huang, Y., Rasco, B.A., 2004. Discrimination of intact and injured *Listeria monocytogenes* by Fourier transform infrared spectroscopy and principal component analysis. J Agric Food Chem 52 (19), 5769-5772

Maquelin, K., Kirschner, C., Choo-Smith, L.P., van den Braak, N., Endtz, H.P., Naumann, D., Puppels, G.J., 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J Microbiol Methods 51 (3), 255-271

Mariey, L., Signolle, J.P., Amiel, C., Travert, J., 2001. Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. Vibr Spectrosc 26 (2), 151-159

Martens, H., Naes, T., 1998 Multivariate Calibration. Wiley and Sons Ltd., Chichester

Naumann, D., 1998. Infrared and NIR Raman spectroscopy in medical microbiology. In: SPIE—International Society for Optical Engineering. Infrared spectroscopy: New tool in medicine. , 1998, Bellingham, Wash, 245–257

Naumann, D., Helm, D., Labischinski, H., 1991a. Microbiological characterizations by FT-IR spectroscopy. Nat 351 81-82

Naumann, D., Helm, D., Labischinski, H., Giesbrecht, P., 1991b. The characterization of microorganisms by Fourier-transform infrared spectroscopy. In: Nelson, W. H. (ed). Modern Techniques for Rapid Microbiological Analysis, John Wiley & Sons New York. pp. 263

Naumann, D., Keller, S., Helm, D., Schultz, C., Schrader, B., 1995. FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the non-invasive characterization of intact microbial cells. J Mol Struct 347 399-405

Ngo Thi, N.A., Kirschner, C., Naumann, D., 2000. FT-IR microspectrometry: a new tool for characterizing micro-organisms. In: SPIE—International Society for Optical Engineering. Biomedical spectroscopy: vibrational spectroscopy and other novel techniques, 2000, Bellingham, Wash, 36-44

Oust, A., Moretro, T., Kirschner, C., Narvhus, J.A., Kohler, A., 2004a. Evaluation of the robustness of FT-IR spectra of *lactobacilli* towards changes in the bacterial growth conditions. FEMS Microbiol Lett 239 (1), 111-116

Oust, A., Moretro, T., Kirschner, C., Narvhus, J.A., Kohler, A., 2004b. FT-IR spectroscopy for identification of closely related lactobacilli. J Microbiol Methods 59 (2), 149-162

Rubio, C., Ott, C., Amiel, C., Dupont-Moral, I., Travert, J., Mariey, L., 2006. Sulfato/thiosulfato reducing bacteria characterization by FT-IR spectroscopy: A new approach to biocorrosion control. J Microbiol Methods 64 (3), 287-296

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least-squares procedures. Anal Chem 36 (8), 1627-1639

WineScan FT120 Type 77110 and 77310 Reference Manual, Issue 4 GB Foss Electric, Denmark, 2001, http://www.foss.dk

Timmins, E.M., Howell, S.A., Alsberg, B.K., Noble, W.C., Goodacre, R., 1998a. Rapid differentiation of closely related *Candida* species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. J Clin Microbiol 36 (2), 367-374

Timmins, E.M., Quain, D.E., Goodacre, R., 1998b. Differentiation of brewing yeast strains by pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. Yeast 14 (10), 885–893

Tintelnot, K., Haase, G., Seibold, M., Bergmann, F., Staemmler, M., Franz, T., Naumann, D., 2000. Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. J Clin Microbiol 38 (4), 1599–1608

Wenning, M., Seiler, H., Scherer, S., 2002. Fourier-Transform Infrared Microspectroscopy, a novel and rapid tool for identification of yeasts. Appl Envir Microbiol 68 (10), 4717–4721

Wolkers, W.F., Oldenhof, H., 2005. In situ FTIR assessment of dried *Lactobacillus bulgaricus*: KBr disk formation affects physical properties. Spect Lett 19 (2), 89-99

Yu, C., Irudayaraj, J., 2005. Spectroscopic characterization of microorganisms by Fourier transform infrared microspectroscopy. BIPMA 77 368–377

Zhao, H., Parry, R.L., Ellis, D.I., Griffith, G.W., Goodacre, R., 2006. The rapid differentiation of *Streptomyces* isolates using Fourier transform infrared spectroscopy. Vibr Spectrosc 40 (2), 213-218

Chapter 5

GENERAL DISCUSSION AND CONCLUSION

5.1 CONCLUSION

The study conducted in Chapter 3 was concerned with the discrimination of wines of three cultivars fermented with different wine yeast strains. It was found that it is indeed possible to discriminate between wines fermented with different yeast stains, using FTIR or GC, for individual sample sets taken during the ageing process of the wines. For the two vintages covered by this study it was observed that the discrimination obtained in each vintage was similar. It was however not possible to see this same discrimination when the data of the two vintages were combined. The discrimination of the fermented wines was found to be similar when using data from GC and FTIR, respectively and analysis with FTIR is considerably faster than analysis by GC. For future studies of similar nature it would be recommended that FTIR is used. It was also possible to discriminate wines fermented with specific tailored genetically modified yeast and wines fermented with non-genetically modified yeasts. This was only performed on a very small sample set and should be further investigated as this is the first study of this nature and differences seen could be as result of the different phenotypes.

Chapter 4 dealt with the study of discriminating wine yeast strains using FTIR and different sample presentations. It was found that it was possible to discriminate between wine yeast strains using FTIR in transmission mode as well as in attenuated total reflectance (ATR) mode. It proved possible to use transmission mode to discriminate between wine yeast strains suspended in different liquid media, water and yeast-peptone-dextrose (YPD). In ATR mode it possible to discriminate between dry yeast samples presented in a granular form, in a pellet form and in a powder form. This highlighted the importance of a standardised protocol for further studies.

5.2 INDUSTRIAL IMPORTANCE

If a need in industry arises that would require the discrimination between wines that was fermented with different wine yeast strains, a calibration must be created that would satisfy all the requirements of the specific need, some of these requirements might include a range of commercial yeasts, some or all major S.A. cultivars, variance-inducing factors, such as maturation of wine, origin of grapes, blending, etc. With a larger study that should include the available commercial GM wine yeast and other GM yeast strains that could potentially be commercialised, the FTIR technique could lead to a possible fast discrimination technique when screening wines for the European Union market to ensure no GM material is used. The establishment of such a calibration will not be trivial and might make such an approach uneconomical.

Through the application of process applied technology (PAT) (McLennan and Kowalski, 1995), possible application in the yeast production process could be to use FTIR technology at-line or even in-line to ensure product purity and consistency. Depending on the stage of the process where the application of the FTIR technology is needed, there might be a requirement for modification of the available instrumentation. Further studies should also be undertaken to ensure that the optimum sample presentation is used for an industrial application.

5.3 LITERATURE CITED

McLennan, F., Kowalski, B.R., 1995 Process analytical chemistry. Chapman & Hall, London