Daniel SCHORN GARCÍA

MONITORING SMALL-SCALE ALCOHOLIC FERMENTATION USING AN ATR-FTIR PORTABLE SPECTROMETER AND MULTIVARIATE ANALYSIS

MASTER THESIS

Supervised by PhD. Montserrat Mestres Solé

Master's Degree in Fermented Beverages

Facultat d'Enologia

UNIVERSITAT ROVIRA I VIRGILI

Tarragona
14th of June 2019

MONITORING SMALL-SCALE ALCOHOLIC FERMENTATION USING AN ATR-FTIR PORTABLE SPECTROMETER AND MULTIVARIATE ANALYSIS

Schorn-García, D., Cavaglia, J., Giussani, B., Boqué, R., Mestres, M.

Dept Química Analitica i Organica, Facultat d'Enologia, Universitat Rovira i Virgilli, 43007 Tarragona, Spain

daniel.schorn@estudiants.urv.cat

Keywords: Process Analytical techniques, mid-infrared, alcoholic fermentation, spoilage, data analysis, process, control

Index

0. Abstract

Fermentation problems during winemaking decrease product quality and may cause important economic losses. Thus, more and more wineries are interested in incorporating quality-by-design strategies instead of postproduction testing. In this sense, early detection with fast analytical techniques, such as FTIR, could be advantageous because they would allow to detect unwanted situations an even 'readjust' the process on time and minimize rejects.

In this study, fermentations under normal conditions and with lactic acid bacteria contamination were monitored with usual classical analytical techniques (density, pH and enzymatic analysis for L-malic acid). A strategy consisting on coupling ATR-FTMIR spectroscopy and multivariate analysis for fermentation process monitoring is proposed.. The aim was to develop a portable, rapid, easy-to-use and economic device to monitor fermentation and to detect deviations.

Multivariate techniques such as exploratory methods (Principal Component Analysis), linear regression methods (Partial Least Squares Regression) and classification methods (Partial Least Squares Discriminant Analysis) were applied in order to control the whole fermentation process. Standard fermentation parameters were successfully predicted, and lactic acid bacteria contamination detected.

La monitorización por metodología clásica de procesos alimentarios, y en concreto en la industria enológica, puede suponer que los posibles problemas que se den durante la fermentación supongan una pérdida de calidad del producto y por tanto una pérdida económica. Por ello, se intenta llevar a cabo hoy en día métodos de control durante el proceso con técnicas analíticas rápidas, como el infrarrojo de transformada de Fourier. Esta metodología de trabajo permite detectar problemas en un estado que pueda ser reajustado sin pérdida de calidad en el producto final.

En el presente estudio, se llevan a cabo dos tipos de fermentaciones, una en condiciones normales y otra con una contaminación por bacterias lácticas. El seguimiento de las fermentaciones se hace tanto por metodología estándar (densidad, pH y análisis enzimático del ácido L-málico) como mediante espectroscopía ATR-FTMIR usando análisis multivariantes de datos. El objetivo es conseguir un equipo portátil, que además sea rápido, económico y fácil de usar para detectar contaminaciones y monitorizar el sistema.

Se utilizan diferentes técnicas multivariantes en el presente estudio, métodos exploratorios, como el Análisis de Componentes Principales; métodos de regresión lineal, como la Regresión por Mínimos Cuadrados Parciales; y métodos de clasificación; como el análisis discriminante por mínimos cuadrados parciales. Todo ello con el objeto de controlar el proceso de fermentación, pero también predecir los parámetros de control básicos y detectar la contaminación por bacterias lácticas.

1. Abbreviations

ATR	Attenuated Total Reflectance
FT	Fourier Transform
LAB	Lactic Acid Bacteria
MIR	Mid Infrared
MLF	Malolactic Fermentation
NIR	Near Infrared
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
PLS	Partial Least Squares
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
SNV	Standard Normal Variate

2. Introduction

Alcoholic fermentation

The basis of the wine production is the yeast biochemical transformation of the must sugars into ethanol. This is the simplest interpretation as carbon dioxide and ethanol are the principal products of the transformation, but the yeasts also produce many other molecules as part of their metabolism. These secondary processes should be also controlled because the compounds produced are responsible for wine complexity, specificity and quality mainly when these are related to organoleptic properties[1].

The main yeast species associated to alcoholic fermentation is *Saccharomyces cerevisiae*. Traditionally, each winemaking region used its native yeasts to carry out spontaneous fermentations. But later, with the industrialization, this process involved inoculating selected yeasts that ensured rapid and complete fermentations, giving rise to very homogenous wines. This is why we are currently facing a moment of change, because today consumers demand unique products, so the current trend is to return to native yeasts to get more organoleptic complexity.

Despite wine production in Spain has a notable impact in the economy and it is present in the country for over thousand years, the industrial development of the sector has been very slow until few years ago. This slow progress is explained because of the traditional concept of wine making as a natural process. Nowadays, it is understood that wine quality needs the implementation of technology, and without it, alcoholic fermentation in wineries would still have a lot of problems related to stuck, sluggish and/or contaminated processes. Stuck and sluggish fermentations are mainly due to deficiency of yeast assimilable nitrogen (the second nutrient yeast requires after carbon sources) or to sudden temperature changes[2]. Spoilage is associated to the presence of other unwanted microorganisms in the must or in unsterile equipment, which find suitable conditions that allow their growth till being a large enough population. Typical microorganisms related to spoilage are acetic acid bacteria but could also be lactic acid bacteria (LAB) and non-*Saccharomyces* yeast.

Lactic acid bacteria are the microorganisms related to malolactic fermentation (MLF) used in winemaking to decrease acidity of wine, typically in red or highly acidic wines. The decrease is explained because of the biochemical transformation of L-malic acid, a dicarboxylic acid, to L-lactic acid, a monocarboxylic acid and carbon dioxide. As it occurs with yeast, the secondary metabolism of the LAB implies the formation of many other compounds that could also affect organoleptic properties of wine[3].

When dealing with red wine, the MLF process is promoted to get a wine deacidification and, therefore, to increase the wine organoleptic quality given that this lower acidity is also more compatible with the high tannicity of these wines. However, in white wines it is mostly considered a drift from the winemaking process that must be avoided because it mainly increases pH and produces the degradation of organic acids such as tartaric, citric or sorbic acids and even results in increased volatile acidity concentrations or toxic compounds[4].

From all these considerations, it is clear that to get a good quality wine a control of the alcoholic fermentation process is needed. In a cellar, the most simple and useful parameters that allow this control are

density (which is related to sugars content), pH (which is related to acids content) and visual and aroma evaluation (which allows to detect abnormal colour and/or aroma). Thus, oenologists use simple instruments and their knowledge to measure these parameters, typically once or twice a day, which is more than enough when the fermentations work correctly[5–7]. However, when a problem is detected, oenologists should act very fast to minimize the effect on the wine quality and, sometimes, the corrective action is not easy nor evident. Thus, instruments able to provide specific real-time information during the course of the fermentation, and not only when the oenologist is able to make a measurement, are required.

The Process Analytical Technologies (PAT) approach is a system for designing, analysing and controlling manufacturing, through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality. The hypothesis behind PAT is that the quality of the products can be incorporated by process design and not by postproduction quality testing. Improved instrumental analytical techniques and multivariate data analysis tools have allowed industries to move analysis closer to the process from off-line measurements via at-line to on-line and in-line analysis, and thus have opened the way for the introduction of the principles of PAT. Therefore, PATs in the wine sector have a great potential to improve product quality and safety, process and resource efficiency, yields, and consequently competitiveness.

PAT was first introduced by the pharmaceutical industries and, like this sector, food has also strict laws and requirements. However, unlike pharmaceutical batches, food matrices are very complex and comprise a wide range of molecules[8]. Furthermore, wine has a great variability along the years because of the climate, harvest time and oenological practices, which introduces a greater complexity when it comes to controlling the process. Thus, to get reliable results, PAT must be optimized for each vineyard or type of wine, because only in this way the wine producer can control thoroughly the winemaking process and achieve precision enology[7].

Even so, different technologies have been satisfactorily applied in wine fermentation processes under the PAT concept. Buratti et al.[1] performed a control of an alcoholic fermentation in wine with four techniques (electronic nose, electronic tongue, mid-infrared and near-infrared spectroscopy) aiming at checking their feasibility to determine the kinetics of the fermentation in comparison with the standard chemical determinations. They concluded that the non-destructive methodologies could be valid and used as standard analysis techniques.

Infrared spectroscopy

Among the different fast techniques, infrared technology seems to be a good option to obtain on-time information[9]. This is because it is a non-destructive technique, which can provide information about molecules or specific functional groups[1] and, overall, it implies a very fast and simple analysis, solvent free and with little or none sample preparation.

The infrared region ranges from 700 to 10^6 nm (IR spectroscopy typically is represented in cm^{-1}, ranging from 10 to 12800 cm^{-1}) of the electromagnetic spectrum and it is divided into three different regions: near, mid and far infrared. Usually near and mid-infrared are used as analytical tools since both allow identification of organic molecules, and far-infrared is restricted to inorganic compounds and organometallic molecules.

The near infrared (NIR) spectrum is characterized by complex overtones and combination bands of fundamental vibrations. Thus, since the useful information is contained in weak bands result of overtones and combinations, the applicability is limited when the concentration of the analyte is low, making difficult to build calibration models and to extract the information of the spectrum.

The mid infrared (MIR) spectrum is the result of fundamental stretching, bending and rotating vibrations, so it can be used to determinate fundamental structural information[10,11]. In general, active fundamental vibration bands have stronger line strength and are better resolved in comparison with near infrared[12]. MIR spectroscopy has the drawback of the high absorbance of OH stretching and bending in aqueous matrices[13]. To get over this problem, attenuated total reflectance (ATR) is employed. This sampling accessory is based on two concepts: first the incident light reaches the sample with an angle of 45º to avoid peak distortions; then the radiation goes through a high-refractive index medium, which is also transparent to IR, such as germanium or diamond crystals[14]. ATR cells simplify sample handling, with a reduction in errors associated to traditional transmission cells.

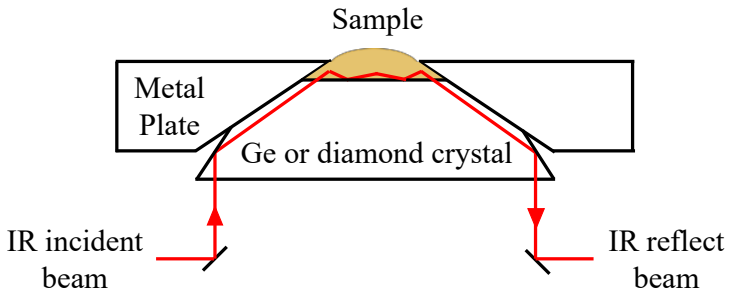

Figure 1. Description of the main parts of the ATR device.

The mechanism of absorption is explained by the rapid change in refractive index across the IR spectrum, and the molar absorption coefficient attributed to the vibrational mode. The IR radiation penetrates in the sample just between 0.5 to 2 μm and, therefore, only one drop of the liquid sample is required. Furthermore, optical path reproducibility enables the possibility of reducing partial spectrum obscurity attributed to water, that is, the saturation of the detector. In the case of hydro-alcoholic matrices such as wine it allows to use the whole range of MIR, what is very useful when dealing with complex mixtures because each molecule contributes in a special way to the overall spectrum[9,15].

In addition, the combined use of ATR with Fourier transform (FT) makes MIR a feasible tool to generate a molecular fingerprint of the sample analysed. The FT-MIR spectrometer parts consist of an IR source, an interferometer, a sampling system (in this case ATR), a detector and finally a data processor. The

source emits an IR beam, which passes through an interferometer. Then the beam is divided by fixed and moving mirrors. As the mirror moves, each wavelength is periodically transmitted or blocked, and then the detector measures the recombined and interfered beam after passing through the sample using a reference laser. The patterns produced by the interferogram are recorded, containing all the acquired information, and resolved in a final spectrum using a FT algorithm[16].

Because of the applicability of this technique to complex matrices, it has been widely applied to food and specifically to wine. Picque et al.[17] in 1993 determined the feasibility of using IR to control the process of alcoholic and malolactic fermentation by the prediction of the principal molecules related to both fermentations. However, MIR instruments were not developed enough at that moment considering that the IR transparent crystals were not common. Afterwards, Teixeira et at.[13] in 2018 compared NIR and MIR with Raman and found all the techniques suitable to determine the same parameters as Picque, and thus showing the growing advance of the infrared technique.

Nowadays, IR spectrometers have not only the forehead advantages (rapid analysis times, small sample sizes, sensitivity, few sample preparation steps, …), but they can also be portable, what allows to move the laboratory to the required place. Equipment is nowadays small, affordable and easy-to-use, with the relative disadvantage of the generation of big amounts of data[18].

Manfredi et al.[19] proved that it is possible to obtain good results based on the IR spectrum obtained with a portable device, providing farmers and certification bodies with a powerful tool to authenticate hazelnuts based on their origin, and reaching classification accuracies of more than 97%. Applicability of portable ATR-FTMIR to wine analysis has recently been proven by Cavaglia et al.[20], by identifying micro-fermentations with a deficiency of yeast assimilable nitrogen.

Multivariate Data Analysis (Chemometrics)

The international Chemometrics Society defines this discipline as the science of relating measurements made on chemical system or process to the state of the system via application of mathematical or statistical methods. Fermentation control with ATR-FTMIR is highly complex as it involves an evolving process which generates a data matrix of three dimensions (cube), $\mathbf{X}_{nxtxm}$, being n the number of samples (batches), m the number of variables (in the present study, wavelengths of the IR spectrum) and t the number of time points collected of the alcoholic or malolactic fermentation. Therefore, a special chemometric data treatment is required to extract the relevant information underneath the whole spectra.

In the present study, a systematic sequence was performed in order to standardize the work when applying different chemometric tools. First, the data obtained with IR spectroscopy requires an unfolding of the three-way matrix. Unfolding can be done in two ways, as ((n x t) x m), so arranging the data sequentially in time; or as (n x (t x m)), where every sample has the spectrum one after another at different time points. The first type of unfolding allows working time by time for all the samples, whereas the second one allows working with the variation in time of the spectrum.

Once the matrix is unfolded, removing irrelevant variation not related to the information of interest, due to equipment and IR source, is required. Typical data pre-processing techniques applied to homogenous aqueous samples in infrared technology are standard normal variate (SNV), which corrects the baseline shift; smoothing, which reduces random noise from the instrumental signal; mean centering, which makes the mean of each variable zero; and first and/or second order derivatives, which may increase differences found in the spectra[21,22].

After data pretreatment, data are ready to be processed and analysed by pattern recognition and/or multivariate calibration algorithms. Pattern recognition techniques can be classified into unsupervised or supervised. Typically, an unsupervised technique is Principal Component Analysis (PCA), which reduces the dimensionality of the original data by preserving most of the information contained. This is achieved by calculating new uncorrelated latent variables, called principal components (PCs), which are linear combinations of the original variables and explain the maximum possible variability in the data in a hierarchical way (PC1 explains more information than PC2, PC2 more than PC3, and so on). When PCA is visually represented it is possible to figure out if there are groups, trends or outlier samples in the dataset analysed.

Unlike PCA, supervised techniques make use of the information about the physicochemical property or the class (category) of the samples analysed and are used for prediction or classification/discrimination purposes. One of the techniques commonly applied is Partial Least Squares (PLS) Regression. Basically, it uses the same concept of PCA, which consists of building new orthogonal (uncorrelated) variables that better explain the variation in the spectra but also correlate with a measured property. Another supervised technique is PLS Discriminant Analysis (PLS-DA), which is based on the PLS regression method, but models a given class/category (arbitrarily assigned a 0 or 1), instead of a measured property.

Finally, to reduce the noise of the IR spectrum, to optimize the extraction of information, and to create more reliable models, with the aim of improving the performance of the pattern recognition or multivariable calibration, it is sometimes useful to select those variables that are most related to the sought information. This is also very useful when using at- or in-line measurements in industries, since it enables the use of non-so high-resolution instruments or to measure just a part of the spectrum[16].

Objectives

The main aim of this study was to determinate the suitability of a portable mid-infrared spectroscopic system, to control the course of the alcoholic fermentation of a white grape must by following the PAT philosophy. Moreover, since the portability of the instrument allows to make measurements in the cellar, the prediction capability of this system was also studied by evaluating parameters useful in the cellar, such as pH or density. Finally, the usefulness of this spectroscopic system to detect the spoilage of LAB (in early stages) was also evaluated.

3. Materials and Methods

Grape must and microorganisms

The must was obtained from Mostos Españoles S.A. (Ciudad Real, Spain) and was collected in September 2017. To avoid natural fermentation, it was stored at -20 ºC in 1.5 L bottles what were filled under N_2 atmosphere. The defrosting process was carried out at 5 ºC and then, it was diluted with MilliQ water to adjust the sugar concentration to 200 ± 10 g/L.

The inoculated yeast used for the experiment was the commercial *S. cerevisiae* active dry yeast "E491" (Vitilevure Albaflor (YSEO), Danstar Ferment A.G., Denmark) and the inoculated bacteria was the commercial freeze-dried blend of *Oenococcus oeni* and *Lactobacillus plantarum* "Co-inoculant Bacteria 3.2" (Anchor Oenology, South Africa).

Micro-fermentations

To have a higher number of monitored samples, the wine-making process was followed at microscale and 6 different batches were monitored. Thus, 34 control fermentations (distributed in 6 different batches) and 33 fermentations with lactic bacteria contamination (distributed in the same 6 different batches) were prepared by using 350 mL of diluted grape must poured into conical flasks. Each flask was supplemented with ENOVIT and ACTIMAX (0.30 g/L each) to get a final concentration of yeast assimilable nitrogen of 300 ± 20 mg/L. All the fermentation processes were carried out at 18ºC.

Each sample was inoculated with 0.105 g of active dry yeast E491 rehydrated in 2 mL for 30 minutes at 35 ºC to get a final concentration of $3 \cdot 10^6$ cell/mL. Related to simulated LAB contamination samples, these were also inoculated in different stages of alcoholic fermentation to reproduce typical critical moments in fermentation. This inoculation of LAB was performed taking into account the producer indications (1 g = 10^{11} cell/mL) to get a final concentration order of 10^6 cell/mL.

An additional experiment (batch 7) was performed by adding known amounts of L-(-)-malic acid (≥95%) (Sigma-Aldrich, Madrid). In this case we followed 3 fermentations under normal conditions and 3 LAB spoilage fermentations that were supplemented with 2 g/L of malic acid (i.e. the samples increased its malic acid content from 1.6 to 3.6 g/L).

Process monitoring

To ensure the correct progress of the fermentation, density and pH were measured twice a day. Density was measured using Densito2Go electronic portable densimeter (Mettler Toledo, United States) and pH using pH 7+DHS pH-meter with a 201 T portable electrode (XS Instruments, Italy). Related to the samples inoculated with LAB, the malolactic fermentation was also followed by measuring L-malic acid with a Y15 Analyser (Biosystems, Barcelona, Spain).

Infrared analysis

Fourier-transform mid-infrared spectra were obtained using a portable 4100 ExoScan FTIR instrument (Agilent, California, USA), equipped with an interchangeable spherical ATR sampling interface, consisting on a diamond crystal window. The spectra were collected in the range of 3999 to 649 cm^{-1}, with a resolution of 8 cm^{-1} and 32 scans at controlled temperature (63 ± 1 ºC). To eliminate interference from laboratory humidity and CO_2 bands the spectra were compensated by running against an air blank.

The samples were collected, at least, once every day to follow both alcoholic and malolactic fermentations until both were finished. To avoid the microorganisms' effect, mainly on the UV absorption when dealing with enzymatic measurements but also on the IR spectra, 1.5 mL of each corresponding sample were centrifuged at 10000 rpm for 10 minutes and the pellet was discarded. One drop of the sample was placed on the crystal and the spectra acquisition was done right after. Each sample was analysed in triplicate with a previous air-background collected spectrum. Before each analysis the crystal was thoroughly cleaned with deionized water and cotton wipes.

Spectra acquisition and multivariate analysis

Spectra acquisition, instrument control and preliminary file manipulation were done using the Microlab PC software (Agilent, California, USA) and data were saved as .spc files. Calculations were carried out using the mean spectra of the triplicates by applying in-house routines written in Matlab v.8.6 R2015b (Mathworks, MA, USA) and the PLS Toolbox software v.6.2 (Eigenvector Research, Manson, WA, USA). After a preliminary exploratory analysis with PCA, PLS regression models were developed to predict fermentation parameters, and PLS-DA models were used to detect LAB spoilage.

4. Results and Discussion

Alcoholic Fermentation

When introducing Process Analytical Technologies by using a multiparametric fast technique in a system, the first step is to find the relationship between the usual or classical process control measurements and the simple at-time measurements with the data obtained with the fast technique (in our study, the ATR-FTMIR technique)[8]. In the case of wine, typical parameters measured in the winery are density and pH.

Figure 2 shows the fermentation kinetics of density and pH in the six batches. Density is an indirect measurement to follow alcoholic fermentation, as it takes into account reducing sugars at the beginning, and in the last hours the ethanol formed in the process. As shown in the evolution curves belonging to six different processes, a similar profile appeared, typical of grape must fermentation[23].

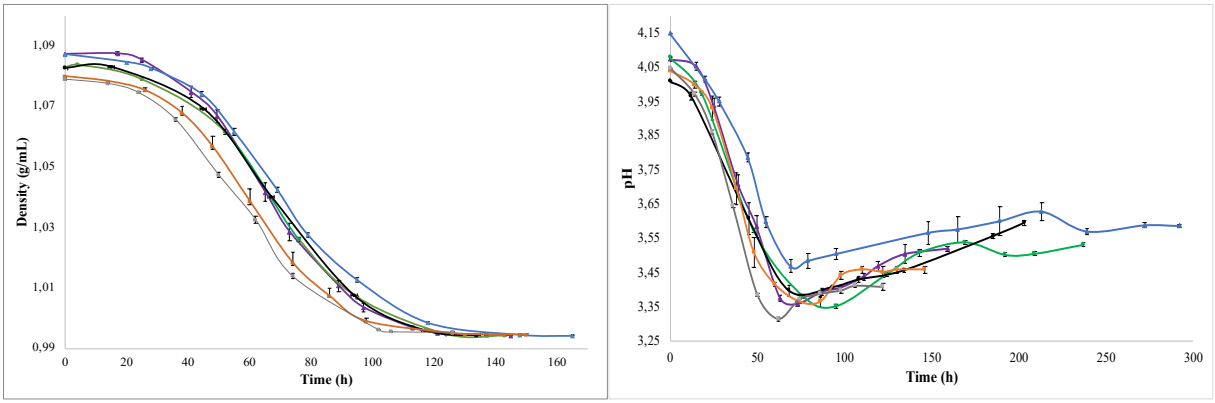

Figure 2. Overall fermentation kinetic profiles of density and pH of the six different batches.

As shown in **Figure 2a**, three different categories for the batches could be distinguished, the first one beginning the fermentation with a density value of 1.087; the second with 1.083 and the last one with 1.080 $g \cdot mL^{-1}$. The different density batches were performed in order to reproduce year-variability in a winery. During the alcoholic fermentation, glucose and fructose were gradually transformed into ethanol by yeasts enzymes until they were totally consumed[24]. At the beginning of the fermentation, the total reducing sugar content ranged between 210 and 190 $g \cdot L^{-1}$. Then, it considerably decreased in the first 80-90 hours (vigorous fermentation) and finally it decreased slowly until total consumption.

As the fermentation process progress, yeasts have various metabolic paths in which they produce different acid compounds such as citric or succinic acids. These compounds, together with the consumption of ammonium, are responsible of the downward trend of pH in fermentation, as it is shown in **Figure 2b**. pH decreased from 4.07 ± 0.08 until it reached its minimum between 60-90 hours, because of the production of acids by yeast. Later, when the malolactic fermentation also occurs (in our study due to LAB spoilage) the pH trend changed and increased because of the transformation of malic acid into lactic acid, so measurement of pH was needed until the end of the MLF. It has to be noted that the control samples do not suffer this process, so these samples are monitored during that time as finished wines. However, since no filtration was applied

(to emulate the LAB samples matrix), we can consider that these samples suffered an "ageing on lees". According to bibliography[25], this kind of ageing improves tartaric acid stability because of the formation of tartaric salts. This fact would explain the pH increase in control fermentations of even 0.2 as the time progressed.

Figure 3 shows the MIR spectra of a single micro-fermentation at different points of the alcoholic fermentation in order to illustrate the changes in the spectrum. It is worth mentioning that the regions from 850 to 649 cm^{-1} were not used in this study due to their high irreproducibility caused by high absorbance values.

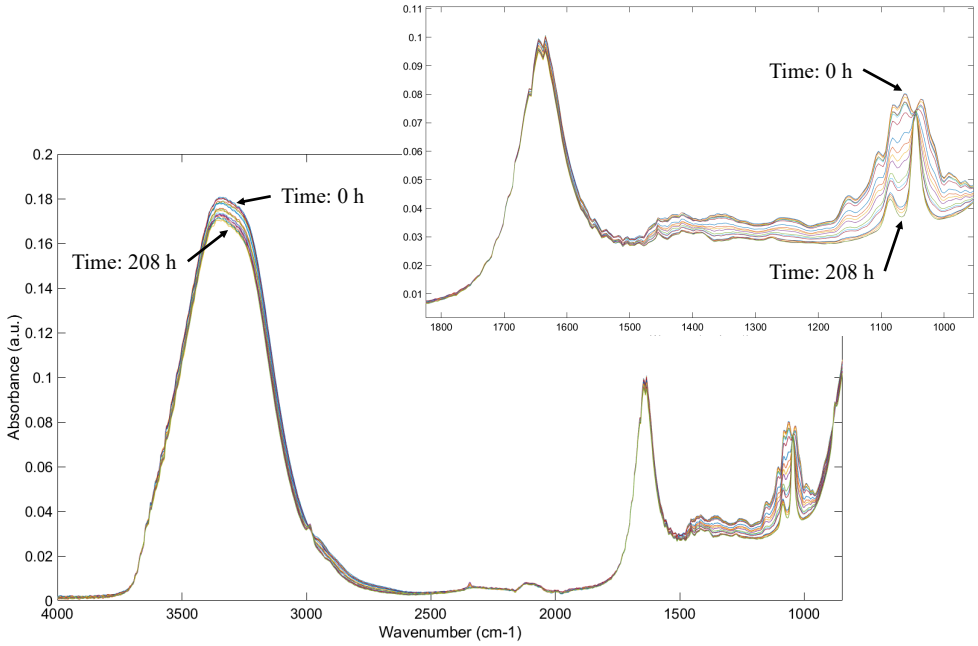

Figure 3. MIR spectrum evolution in a typical grape must alcoholic fermentation, from 0h to 208h

Grape must, wine and all the intermediate states are complex mixtures that share the same molecules or even the same functional groups in different molecules, so it is very difficult to assign specific peaks to a given compound. However, it is reasonable to attribute the largest peaks to chemical bond vibrations of the principal compounds[6]. Several authors[26–29] had assigned the region from 3700 to 3000 cm^{-1} to water because of the OH stretching absorption, which is also related to the band between 1780 and 1500 cm^{-1}. As water composition does not change during alcoholic fermentation, the changes observed in these regions correspond to the contribution of ethanol (when increasing throughout the alcoholic fermentation) over the OH stretching signal and to the pH variations that affect water bonds.

As shown in **Figure 3** the region from 1500 to 900 cm^{-1} has the largest variability, what is expected because acids, alcohols and sugars present the major peak absorptions in that area. The complex bands between 1500 and 1100 cm^{-1} are associated to CH_2, C-C-H and H-C-O typical functional groups of organic acids, alcohols and proteins, but they can also be found in sugars or other molecules. The major variation from 1100 to 900 cm^{-1} is related to C-C and C-O stretching of sugars, glucose and fructose, which are consumed due to

the biochemical yeast transformation to ethanol. The peak around 1042 cm^{-1} inside this major area could be assigned to C-O stretching of alcohols, mainly ethanol[30].

Finally, other minor changes in the spectra are found around 3000 and 2800 cm^{-1}, and also around 2341 cm^{-1}. The first area has been related to stretching of CH_3 and CH_2 and considering that it is only visible while fermentation progress, some authors assigned it to ethanol[5,27,28]. The peak around 2341 cm^{-1} is only visible in fermenting must, as it is related to CO_2 absorbance band[5,28].

Preliminary exploratory analysis with PCA was performed by using the mean spectra in order to check the repeatability of the measurements, to detect outliers, and to recognize patterns in the samples' distribution. PCA sample scores were calculated on PC1 and PC2 and these were plotted versus time in order to find time-related changes that occur during grape must fermentation (**Figure 4**).

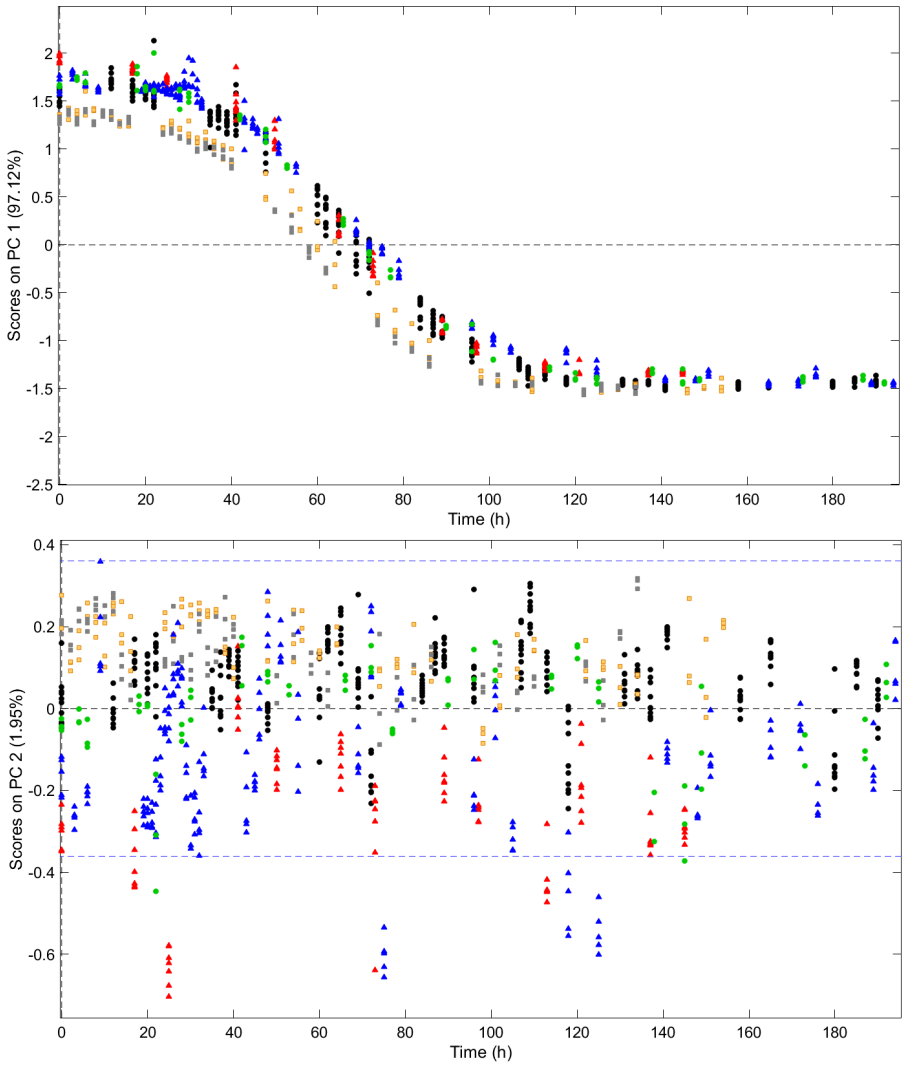

Figure 4. Score plots obtained from MIR spectra in the area defined by a) the PC1 vs time (h); and b) the PC2 vs time (h)

As **Figure 4** illustrates, the scores from the first PC showed a similar trend to must density evolution. This behaviour highlighted that this PC explains alcoholic fermentation as the decrease of sugars. This is

confirmed with the loadings of the variables in the first PC (**Figure 1 - Annex I**), which shows the importance of the region from 1700 to 950 cm^{-1}, related to C-O and C-H of ethanol and sugars.

In the second PC, a trend is observed between batches as samples from the same batch are grouped around a certain interval of PC2. Therefore, PC2 is able to detect the differences between every batch. As every batch had a different sugar concentration and were prepared in six different moments it is assumed that this variability is possible. Nevertheless, the information contained in PC1 is the 97.12% of the information contained in the spectra and PC2 only included a 1.95 % of the spectroscopic information.

To check the possibility to predict the fermentation status, a PLS regression between the spectra acquired through fermentation (hours elapsed since yeast inoculation, ranging from 20 to 120 hours) and the predicted time was studied. As it can be seen in **Figure 4,** we only considered this time lapse as the most vigorous step of fermentation (also called tumultuous fermentation) because at this moment is when the changes are really important. In fact, before and after this step it was very difficult to detect differences in the spectra. **Figure 2 – Annex I**, shows the PLS regression model for time prediction, which was cross-validated (random samples, 15 splits and 10 iterations). The resulting model had a linear regression of R^2 of 0.971, and a root mean square error of cross-validation (RMSECV) of 5.16 hours, and no cross-validation bias was detected. It must be taken into account that batch variability was introduced in the model as it was built with all batches. So, the model is able to predict within 12 hours at what moment of the process the fermenting must is, without year-variability correction. From the results obtained, we conclude that the portable FTIR used allows monitoring the alcoholic fermentation with the possibility to know at what moment of this process we are.

PLS Regression

The use multivariate analysis (chemometrics) is required to extract the maximum useful information from the data. The PLS regression method was used to build four different multivariate calibration models between the FT-MIR spectra and the reference values of density, pH and L-malic acid (during alcoholic fermentation and just in final wine). The spectral pre-processing used was evaluated based on the lowest RMSECV (root mean square error of cross-validation) values of the models and was optimized for every parameter. Different second-order smoothing filters were tried, by adjusting a polynomial to different spectrum points (ranging from 7 to 15 points). Finally, the pre-processing techniques selected were the ones indicated in **Table 1**.

The number of optimal PLS factors in the final models was obtained by using random subsets of samples. In that way, the samples were divided into fifteen random splits, and the model was built with the remaining ones. Then, the model was used to predict the left-out samples. The procedure was repeated until all the groups were excluded. Finally, the procedure was repeated with ten different splits of the samples in order to get a better model, and the average of the ten models was used as the final model. The optimal number of PLS factors was the one which provided the best (lowest) RMSECV.

Table 1. Analytical Performance Parameters of the Multivariate Calibration Models built by PLS Regression

Dependent variable	Variable range	Nº samples	Pre-processing (Variable Selection)	PLS factors	Calibration		Cross-validation		Prediction		
					RMSEC	R^2	RMSECV	R^2	Nº Samples	RMSEP	R^2
Density	0.994 - 1.087	522	Smooth[a] - SNV Mean Centre (No)	4	0.0011	0.9990	0.0012	0.9989	580	0.0013	0.9986
pH	3.32 – 4.15	352	Smooth[b] – SNV Mean Centre (Yes[c])	6	0.06	0.9259	0.06	0.9259	376	0.06	0.9345
L-malic acid	0.05 – 2.00	237	Smooth[b] - SNV Mean Centre (Yes[d])	3	0.21	0.9024	0.22	0.8968	267	0.21	0.9106
L-malic acid*	0.05 – 2.00	74	Smooth[b] - SNV Mean Centre (Yes[d])	3	0.18	0.9212	0.19	0.9063	84	0.20	0.9074

RMSEC: Root mean square error of calibration; RMSECV: root mean square error of cross-validation; RMSEP: root mean square error of prediction; R^2: coefficient of determination; [a]Second-order smoothing polynomial through seven points; [b]Second-order smoothing polynomial through fifteen points; [c]Selectivity ratio algorithm used; [d]Manual variable selection to regions more related to organic acids dec6(1320 – 1109 cm^{-1}); *The model was performed with wine (density<0.995 $g \cdot L^{-1}$)

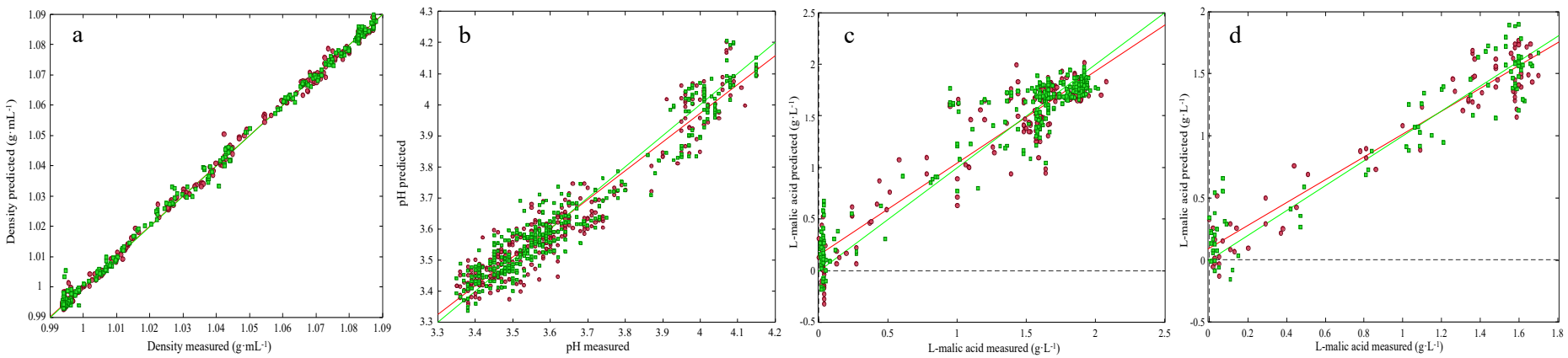

Figure 5. Scatter plots of reference measurements versus MIR predictions for density (a), pH (b), L-malic acid (c) and L-malic acid in wine (d). ● Calibration and ■ Validation

To evaluate the ability of the model to describe future samples, at the beginning the samples were divided approximately into two halves: one half to build the model and the other half as a validation set. The prediction ability of the model was evaluated with the RMSEP (root mean square error of prediction) calculated for the validation set.

Density

As it was shown in **Figure 4**, the first PC against the time displayed a shape similar to the density evolution shown in **Figure 2**; therefore, it was likely that the PLS regression shows a great ability to predict density. The reference measurement equipment has an error of 0.0010 $g \cdot mL^{-1}$; and the RMSECV and RMSEP were 0.0012 and 0.0013 $g \cdot mL^{-1}$, respectively. The wide range of density values determined all along alcoholic fermentation, together with the high number of samples analysed allowed building a good model for a parameter with high variability, such as density.

It has to be noted that density prediction during alcoholic fermentation had been already performed by Fernández-Novales, et al.[31] using a fiber NIR and the cross-validated value obtained was 0.0004 $g \cdot mL^{-1}$ (RMSECV). Although this prediction error is lower that the obtained with our portable FTIR device, it must be taken into account that the number of batches of that experiment was just one, so the variability considered is very low.

pH

pH is a parameter that has a strong dependency on the different compounds of the sample and, in the case of wine, mainly on organic acids but also ethanol. The models obtained for pH prediction, showed very good results as the R^2_{CV} and R^2_{Pred} had a value over 0.90. It seems logical that the PLS model needs 6 factors, as many compounds have an influence on pH (compared with density) so it was necessary to perform a variable selection to achieve better results. The regions selected in the model were 3700 to 2900 and 1750 to 850 cm^{-1}, which are related to water, organic acids, and sugars. This selection is due to the fact that pH changes in the matrices affect not only the acids.

Predicting pH with mid-infrared spectroscopy had been already studied in grape must and wine, with a performance very similar to the results obtained in the present study. Thus, Teixeira et al.[13] obtained a RMSEP of 0.05 in wine and Ozturk et al.[32] performed a model with a result of prediction of 0.09. On the other hand Shah et al.[14], obtained an error of cross-validation in grape must of 0.07. From these results, it can be concluded that the results obtained for pH prediction with a portable MIR are similar to the ones obtained with a benchmark one, during the alcoholic fermentation process and the "ageing on lees".

L-malic acid

L-malic acid was determined during the alcoholic fermentation since LAB and yeasts were co-inoculated; and LAB could begin their metabolism.

The first multivariate model built showed that the downward signal of sugars was used to explain also the decrease of L-malic acid before the end of the alcoholic fermentation, but the model was not able to establish a good correlation. This behaviour was the one expected because there is no metabolic relationship between these compounds. Therefore, it was necessary to improve the performance of the model by selecting specific spectral regions, which really were related to the malolactic fermentation process. That selection was firstly made by using algorithms that the PLS Toolbox software proposed. However, the algorithms selected different variables in each run and always selected the sugar region as the one necessary to establish the correlation. This fact demonstrated that the algorithms implemented in the software were not able to find the variables that really explain the malolactic process. The next step was to manually select the variables that bibliography describes as the most important for organic acids, specially L-malic and L-lactic acids[30,33].

The prediction results obtained appear not so accurate as density and pH; however, they PLS models need less factors to explain the correlation between L-malic acid and the MIR spectra. It is expected that a compound whose variation ranges from 0.00 (below LOD) to 2.00 $g \cdot L^{-1}$ and being in a concentration less than 1% in the must would have a high error. The error of prediction (RMSEP) was 0.21 $g \cdot L^{-1}$ whereas Urtubia et al.[34], with a benchtop FT-IR instrument, obtained an error of 0.34 $g \cdot L^{-1}$ for Cabernet Sauvignon and of 0.29 $g \cdot L^{-1}$ for the rest of varieties studied. This shows that a model built for a molecule with a very low concentration with a portable ATR-FTMIR is comparable to a model built with a benchtop IR instrument.

To evaluate the capacity of the model to predict L-malic acid concentration without the interference of sugars, a model was built with just wine (density >0.995 $g \cdot mL^{-1}$) to check the influence of the alcoholic fermentation in the model. The prediction error achieved seems to be a little more accurate (0.20 $g \cdot L^{-1}$). Compared to other studies such the ones performed by Patz et al.[35], the RMSEP value of 0.63 $g \cdot L^{-1}$ obtained with a benchtop mid-infrared spectrometer showed that variable selection is a powerful tool to reduce non-relevant information of the spectrum and obtain better results.

LAB spoilage

In this study, an early stage deviation was intentionally promoted by adding LAB to an alcoholic fermentation process with the purpose of detecting the contamination even at the first moments of the process. Two problems were detected in this approach. First, as the grape must is a high sugar content matrix, it was difficult to see variations of L-malic acid concentrations, which is present initially at a maximum 2 g/L. On the other hand, as it was reported in the literature[36], yeasts are also able to use malic acid as a part of their metabolism. So, the attempts to detect contamination at early stages in grape must were unsuccessful. It was found that the moment in which the detection could be accomplished is when the yeasts have consumed all the sugar, that is, in final wine.

As it was concluded before (PLS regression) the spectroscopic regions related to organic acids (L-malic and L-lactic acid) are the most promising to detect malolactic fermentation problems, as it was demonstrated that it is possible to build models with good correlation with the amount of L-malic acid. The

consumption of L-malic acid could be considered as the main difference between a normal and a contaminated process as it is the reagent or product with the highest concentration in MLF.

Preliminary exploratory analysis with PCA was performed on the mean spectra in order to recognize patterns in the samples' distribution. The preprocessing used was second-order smoothing polynomial through fifteen points with first derivative, SNV and mean centering. Then the sample scores were calculated on PC1 and PC2 and were plotted against each other in **Figure 6**.

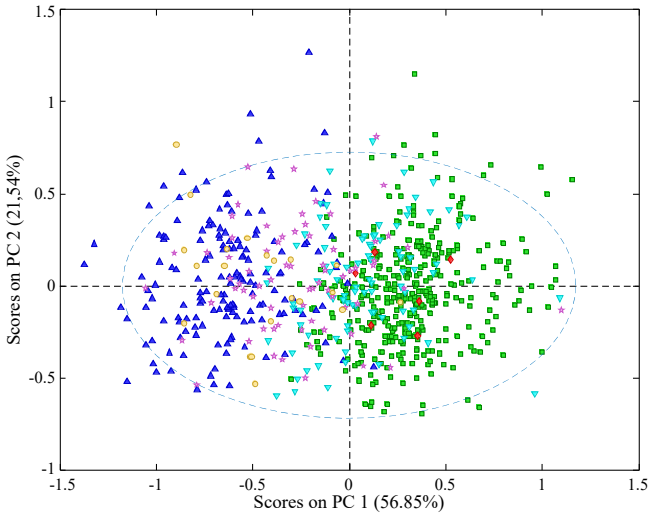

Figure 6. Score plot of PC1 (43.19 %) vs PC2 (24.03 %) indicating the percentage of malolactic fermentation: 0% (green), 0-25 % (red), 25-50 % (light blue), 50-75 % (pink), 75-100 % (yellow) and 100 % (dark blue)

As **Figure 6** shows, a trend is detected in PC1 as the malolactic fermentation moves forward. This behaviour highlighted that this PC explains malolactic fermentation as the downward of L-malic acid. In the PC2, a high variability is observed as the result of introducing batch differences in the model (every batch had a different sugar concentration). In order to perform better malolactic fermentation detection, supervised discrimination models (PLS-DA) were built with the intention of focusing on variability related to detect LAB contamination (**Figure 7**). It was assigned 1 to control fermentations and 0 to the BAL spoilage fermentations.

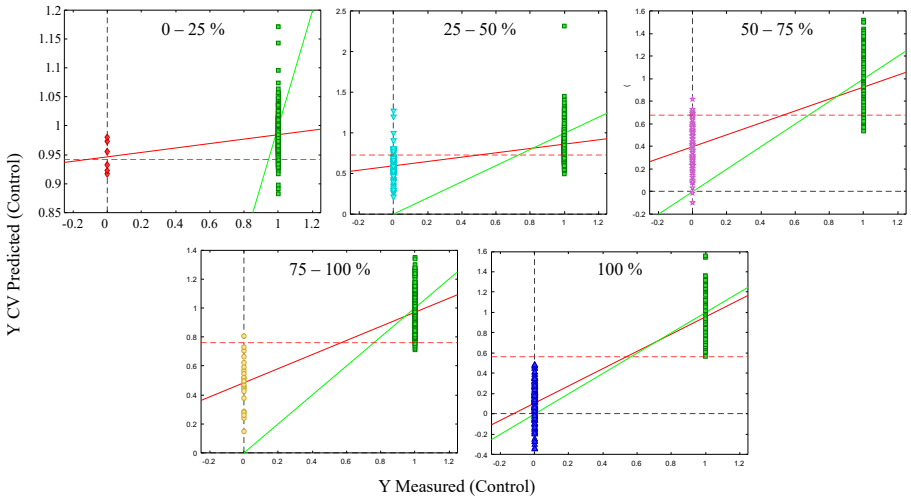

Figure 7. Discrimination plots between normal fermentation (1, green) and different progress steps of malolactic fermentation (0).

As **Figure 7** shows, the ability to recognize different states of LAB contamination increased as malolactic fermentation advanced. It can be seen that the threshold, that is the red discontinuous line, dismissed in every PLS-DA as the differences between deviated and control fermentation increased. From 0 to 50% it was difficult to distinguish between control and deviated processes. This is explained first because of the lack of points between 0 and 25%. Secondly, the differences between the normal and deviated fermentations are slight as much in L-malic acid as in pH differences. 50-75% had an improved trend showing a good differentiation. Finally, 75 to 100% of malolactic advance was well differentiated, until the 100% of malolactic fermentation when this differentiation was total.

5. Conclusions and Future perspectives

Conclusions

The drawn conclusions of this study about monitoring of grape must alcoholic fermentation using a portable ATR-FT-MIR are:

- The kinetic evolution of the grape must fermentation obtained with the PC1 versus time showed a similar trend with respect to the density curve.
- ATR-FT-MIR is a reliable tool to predict standard fermentation control parameters (density, pH and L-malic acid).
- PLS-DA models are able to detect LAB contamination before the end of malolactic fermentation and are able to distinguish between the different stages of the contamination.
- A fast, simple and at-line analytical device has a great potential to monitor alcoholic fermentation and offers the possibility to eventually correct the process.

Future perspectives

The prediction error of the **PLS regression** models for pH and malic acid and for the **LAB spoilage** at first steps of the process, which had slight differences with control fermentation, made difficult to detect LAB contamination before half of the deviation had happened. As the initial L-malic acid concentration was low (2 g/L), it is possible that an increase of concentration would allow to improve the detection of the deviation. The dataset 7 (3 normal and 3 LAB contaminated fermentations) was supplemented with L-malic acid until 3.6 g/L, which is still a normal concentration in wines. This allowed to enlarge the variable range in L-malic acid prediction as it is shown in **Figure 8**.

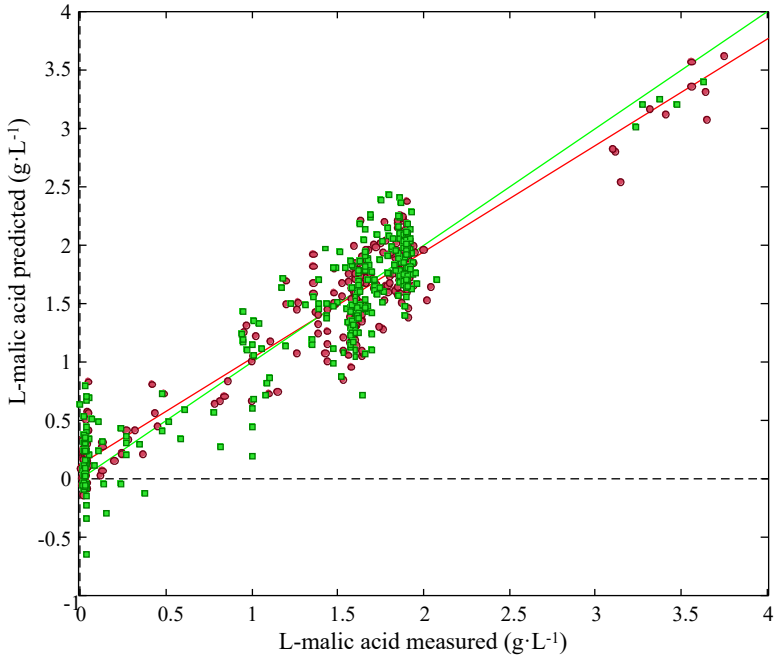

Figure 8. Scatter plots of reference measurements and MIR predictions for L-malic acid. ● Calibration and ■ Validation

The error of prediction achieved when increasing L-malic acid concentration was 0.29 $g \cdot L^{-1}$, showing higher error in the enlarged variable range. This seems logical as it existed a gap between the previous model and the added samples. On the other hand, the model prediction ability remains linear at higher concentrations, so it would allow us to build a better prediction model with more samples.

An exploratory analysis with PCA was performed on the mean spectra of wines with a higher amount of L-malic acid in order to recognize patterns in the samples' distribution, as it is shown in **Figure 6**. The preprocessing used was the same, second-order smoothing polynomial through fifteen points with first derivative, SNV and mean centering. Then the sample scores were calculated on PC1 and PC2 and were plotted against each other (**Figure 9**).

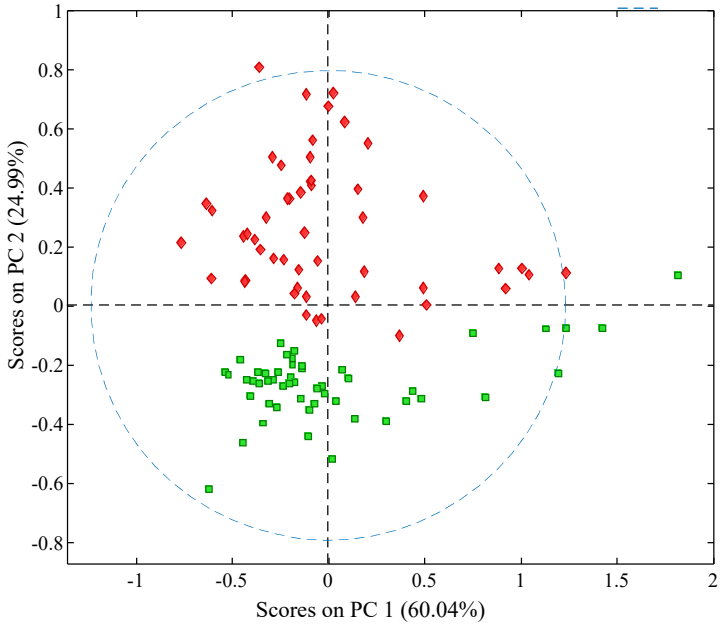

Figure 9. Score plots of PC1 (60.04 %) vs PC2 (24.99 %) indicating control (green) or BAL contaminated (red) fermentations.

As it can be concluded from **Figure 9**, two different groups were observed in PC2. In this case, final alcoholic fermentation in LAB contaminated process remains with a L-malic acid concentration of 2.5 $g \cdot L^{-1}$, meaning a 70% of the initial content. A higher concentration of a molecule improves infrared signal resolution and the model is then able to detect slighter differences.

Since L-malic acid concentrations in wines are up to 8 $g \cdot L^{-1}$ in cooler areas, this work has a wide range to improve infrared prediction and to detect LAB spoilage in wines with higher amount of L-malic acid. So, future work will focus on building predictive models of L-malic acid up to 8 $g \cdot L^{-1}$, thus covering the typical content; and to evaluate different concentrations of LAB spoilage as they can affect malolactic fermentation kinetics.

6. Acknowledgements

En primer lugar, me gustaría agradecer a Montse y Laura, porque desde el primer momento he podido contar con su ayuda para llevar adelante este TFM. Estoy muy agradecido a Laura por sus consejos y por tener siempre un momento para escucharme. A Montse quiero agradecer sus increíbles ideas, su capacidad de ver el lado bueno de todo y de mantenernos animados en todo momento. **Gracias, Montse y Laura.**

Durante todo el trabajo, he hecho uso de una herramienta que se ha convertido poco a poco en un pilar muy importante en mi desarrollo científico, la quimiometría, y no puedo olvidarme de las personas tan capaces y que han sabido transmitirme parte de su conocimiento de una manera tan clara y con tanta paciencia. A Ricard por su implicación en el TFM y por tener siempre un momento para las dudas. A Bárbara no puedo estar más agradecido por hacer que casi 1000 km no sean impedimento para las mejores clases de quimiometría, y su increíble implicación en este proyecto. **Gracias, Ricard y Bárbara.**

No puedo olvidar a Julieta por su acogida desde el primer momento, por alegrar nuestro tiempo en el laboratorio, por su incansable trabajo con nuestras fermentaciones y por ser el pilar más grande en el que apoyarme para empezar a aprender todo lo relacionado con el TFM. **Gracias, Julieta.** Dentro del laboratorio no puedo olvidarme de agradecer la sonrisa de Carolina y su apoyo durante todo el desarrollo del proyecto; a Aziz por amenizar mi tiempo en el laboratorio y acercarme al mundo del aceite. **Gracias, Carolina y Aziz.**

A mi nueva amiga y futura compañera de piso, Candela, no puedo estar más agradecido de la casualidad que ha sido vivir en el mismo edificio, porque me has dado tantos buenos momentos durante el TFM y me has apoyado de una manera casi incansable. Espero poder seguir contando contigo durante los próximos años porque quiero devolverte todo lo que has hecho por mí, tanto con respecto a las clases como fuera. **Gracias, Candela.**

Además, gracias a ella he conocido a dos personas maravillosas, Aitor y Mari, que me han acogido en el departamento de Bioquímica, desde el momento de empezar a usar el Miura, yendo al gimnasio juntos o a las tardes de cañas, aconsejándome tanto sobre el TFM. **Gracias, Aitor y Mari.**

Los nuevos amigos se unen a los que ya estaban, a los que se van uniendo en cada etapa de mi vida y se mantienen incansablemente en ella, preguntándome porque tengo que ir en fin de semana al laboratorio o a horas tan intempestivas, si solo soy "un químico que ha encontrado la excusa perfecta para beber vino". **Gracias, Rubén, Rebeca, Iris, Laura y Keren. Gracias, Cris, Bea y Pau.**

Finalmente, no puedo expresar el gran agradecimiento que tengo a las dos mujeres más importantes de mi vida, mi madre y mi abuela. Ya que, aunque no entiendan totalmente mi trabajo en el laboratorio, me preguntan y me dejan que me explaye todo lo que quiera sobre él, para finalmente preguntarme si estoy contento con lo que hago y ser felices sabiendo que es así. Gracias por haberme hecho llegar hasta aquí. **Gracias, mamá y abuela.**

"*En la investigación es incluso más importante el proceso que el logro mismo*" (E. Muñoz)

7. References

1. Buratti, S., Ballabio, D., Giovanelli, G., Dominguez, C. M. Z., Moles, A., Benedetti, S., & Sinelli, N. (2011). Monitoring of alcoholic fermentation using near infrared and mid infrared spectroscopies combined with electronic nose and electronic tongue. *Analytica Chimica Acta*, *697*, 67–74.

2. Hernández, G., León, R., & Urtubia, A. (2016). Detection of abnormal processes of wine fermentation by support vector machines. *Cluster Computing*, *19*, 1219–1225.

3. Cozzolino, D., Mccarthy, J., & Bartowsky, E. (2012). Comparison of near infrared and mid infrared spectroscopy to discriminate between wines produced by different Oenococcus Oeni strains after malolactic fermentation: A feasibility study. *Food Control*, *26*, 81–87.

4. Carreté i Aixalá, R. (2002). Influència de components dels vins sobre el metabolisme d'Oenococcus oeni i sobre la seva dinàmica de poblacions a la fermentació malolàctica (Doctoral Thesis). Universitat Rovira i Virgili. Facultat d'Enologia de Tarragona, Tarragona.

5. Wu, Z., Xu, E., Long, J., Zhang, Y., Wang, F., Xu, X., Jin, Z., & Jiao, A. (2015). Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy. *Food Control*, *50*, 405–412.

6. Urtubia, A., Pérez-correa, J. R., Pizarro, F., & Agosin, E. (2008). Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. *Food Control*, *19*, 382–388.

7. Cozzolino, D. (2016). State-of-the-art advantages and drawbacks on the application of vibrational spectroscopy to monitor alcoholic fermentation (beer and wine). *Applied Spectroscopy Reviews*, *51*, 282–297.

8. Van Den Berg, F., Lyndgaard, C. B., Sørensen, K. M., & Engelsen, S. B. (2013). Process Analytical Technology in the food industry. *Trends in Food Science & Technology*, *31*, 27–35.

9. Roychoudhury, P., Harvey, L. M., & McNeil, B. (2006). The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Analytica Chimica Acta*, *571*, 159–166.

10. dos Santos, C. A. T., Páscoa, R. N. M. J., & Lopes, J. A. (2017). A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle. *Trends in Analytical Chemistry*, *88*, 100–118.

11. Schalk, R., Geoerg, D., Staubach, J., Raedle, M., Methner, F.-J., & Beuermann, T. (2017). Evaluation of a newly developed mid-infrared sensor for real-time monitoring of yeast fermentations. *Journal of Bioscience and Bioengineering*, *123*, 651–657.

12. Roberts, J. J., Power, A., Chapman, J., Chandra, S., & Cozzolino, D. (2018). Vibrational Spectroscopy Methods for Agro-Food Product Analysis. In Lopes, J. & Sousa, C. (Eds.), *Handbook: Comprehensive*

Analytical Chemistry (pp 51–68). Amsterdam: Elsevier.

13. Teixeira dos Santos, C. A., Páscoa, R. N. M. J., Porto, P. A. L. S., Cerdeira, A. L., González-Sáiz, J. M., Pizarro, C., & Lopes, J. A. (2018). Raman spectroscopy for wine analyses: A comparison with near and mid infrared spectroscopy. *Talanta*, *186*, 306–314.

14. Shah, N., Cynkar, W., Smith, P., & Cozzolino, D. (2010). Use of attenuated total reflectance midinfrared for rapid and real-time analysis of compositional parameters in commercial white grape juice. *Journal of Agricultural and Food Chemistry*, *58*, 3279–3283.

15. Pomerantsev, A. L., & Ye Rodionova, O. (2012). Process analytical technology: a critical view of the chemometricians. *Journal of Chemometrics*, *26*, 299–310.

16. Borràs Iglesias, E. (2016). Instrumental and chemometric methodologies to assess sensory quality of mediterranean food (Doctoral Thesis). Universitat Rovira i Virgili. Facultat d'Enologia de Tarragona, Tarragona.

17. Picque, D., Lefier, D., Grappin, R., & Corrieu, G. (1993). Monitoring of fermentation by infrared spectrometry: Alcoholic and lactic fermentations. *Analytica Chimica Acta*, *279*, 67–72.

18. Crocombe, R. A. (2018). Portable Spectroscopy. *Applied Spectroscopy*, *72*, 1701–1751.

19. Manfredi, M., Robotti, E., Quasso, F., Mazzucco, E., Calabrese, G., & Marengo, E. (2017). Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *189*, 427–435.

20. Cavaglia, J., Giussani, B., Mestres, M., Puxeu, M., Busto, O., Ferré, J., & Boqué, R. Early detection of undesirable deviations in must fermentation using a portable FTIR-ATR instrument and multivariate analysis. *Journal of Chemometrics*.

21. Vera Carrasco, L. (2010). Sensory Quality Control of Alcoholic Beverages Using Fast Chemical Sensors (Doctoral Thesis). Universitat Rovira i Virgili. Facultat d'Enologia de Tarragona, Tarragona.

22. Rinnan, A., Van Den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, *28*, 1201–1222.

23. Emparán, M., Simpson, R., Almonacid, S., Teixeira, A., & Urtubia, A. (2012). Early recognition of problematic wine fermentations through multivariate data analyses. *Food Control*, *27*, 248–253.

24. Di Egidio, V., Sinelli, N., Giovanelli, G., Moles, A., & Casiraghi, E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *European Food Research and Technology*, *230*, 947–955.

25. Charpentier, C. (2010). Ageing on lees (sur lies) and the use of speciality inactive yeasts during wine fermentation. In Reynolds, A. G. (Ed.), *Managing Wine Quality: Oenology and Wine Quality* (pp 164–187). Cambridge: Woodhead.

26. Condurso, C., Cincotta, F., Tripodi, G., & Verzera, A. (2018). Characterization and ageing monitoring of Marsala dessert wines by a rapid FTIR-ATR method coupled with multivariate analysis. *European Food Research and Technology*, *244*, 1073–1081.

27. Di Egidio, V., Sinelli, N., Giovanelli, G., Moles, A., & Casiraghi, E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *European Food Research and Technology*, *230*, 947–955.

28. Urtubia, A., Ricardo Pérez-Correa, J., Pizarro, F., & Agosin, E. (2008). Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. *Food Control*, *19*, 382–388.

29. Wu, Z., Long, J., Xu, E., Wu, C., Wang, F., Xu, X., Jin, Z., & Jiao, A. (2015). Application of FT-NIR spectroscopy and FT-IR spectroscopy to Chinese rice wine for rapid determination of fermentation process parameters. *Analytical Methods*, *7*, 2726–2737.

30. Pretsch, E., Clerc, T., Seibl, J., & Simon, W. (1991). Tablas para la elucidación estructural de compuestos orgánicos por métodos espectroscópicos. Madrid: Alhambra Longman, S.A.

31. Fernández-Novales, J., López, M. I., Sánchez, M. T., García, J. A., & Morales, J. (2008). A feasibility study on the use of a miniature fiber optic NIR spectrometer for the prediction of volumic mass and reducing sugars in white wine fermentations. *Journal of Food Engineering*, *89*, 325–329.

32. Ozturk, B., Yucesoy, D., & Ozen, B. (2012). Application of Mid-infrared Spectroscopy for the Measurement of Several Quality Parameters of Alcoholic Beverages, Wine and Raki. *Food Analytical Methods*, *5*, 1435–1442.

33. Bureau, S., Cozzolino, D., & Clark, C. J. (2019). Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biology and Technology*, *148*, 1–14.

34. Urtubia, A., Ricardo Pérez-Correa, J., Meurens, M., & Agosin, E. (2004). Monitoring large scale wine fermentations with infrared spectroscopy. *Talanta*, *64*, 778–784.

35. Patz, C. D., Blieke, A., Ristow, R., & Dietrich, H. (2004). Application of FT-MIR spectrometry in wine analysis. *Analytica Chimica Acta*, *513*, 81–89.

36. Son, H. S., Hwang, G. S., Park, W. M., Hong, Y. S., & Lee, C. H. (2009). Metabolomic characterization of malolactic fermentation and fermentative behaviors of wine yeasts in grape wine. *Journal of Agricultural and Food Chemistry*, *57*, 4801–4809.

8. Annex 1

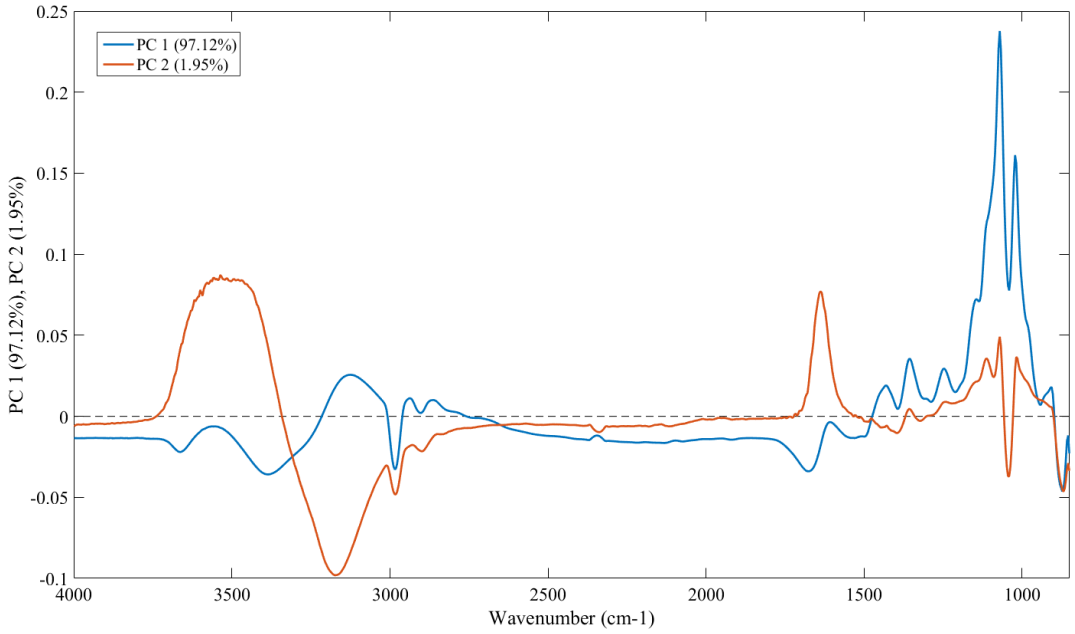

Figure 1. Loadings of the PC1 and PC2 of the PCA of the six batches.

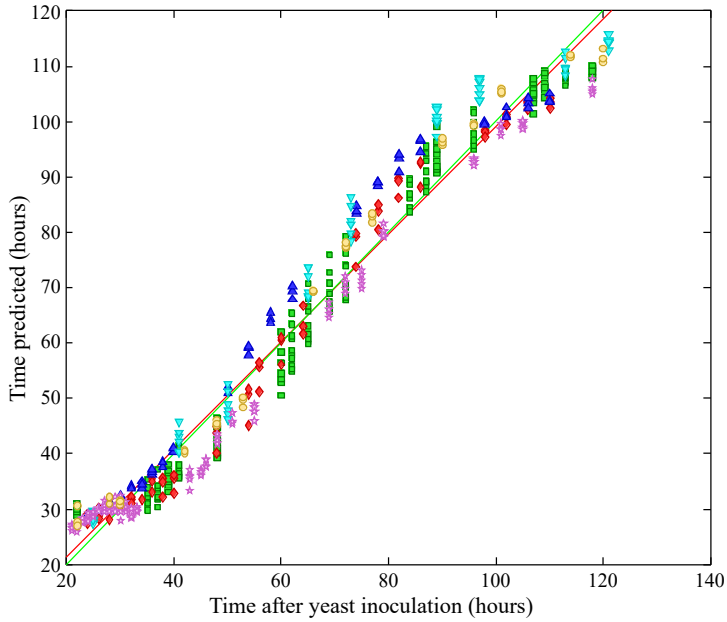

Figure 2. Scatter plots of time after yeast inoculation and MIR predictions of different fermentations.