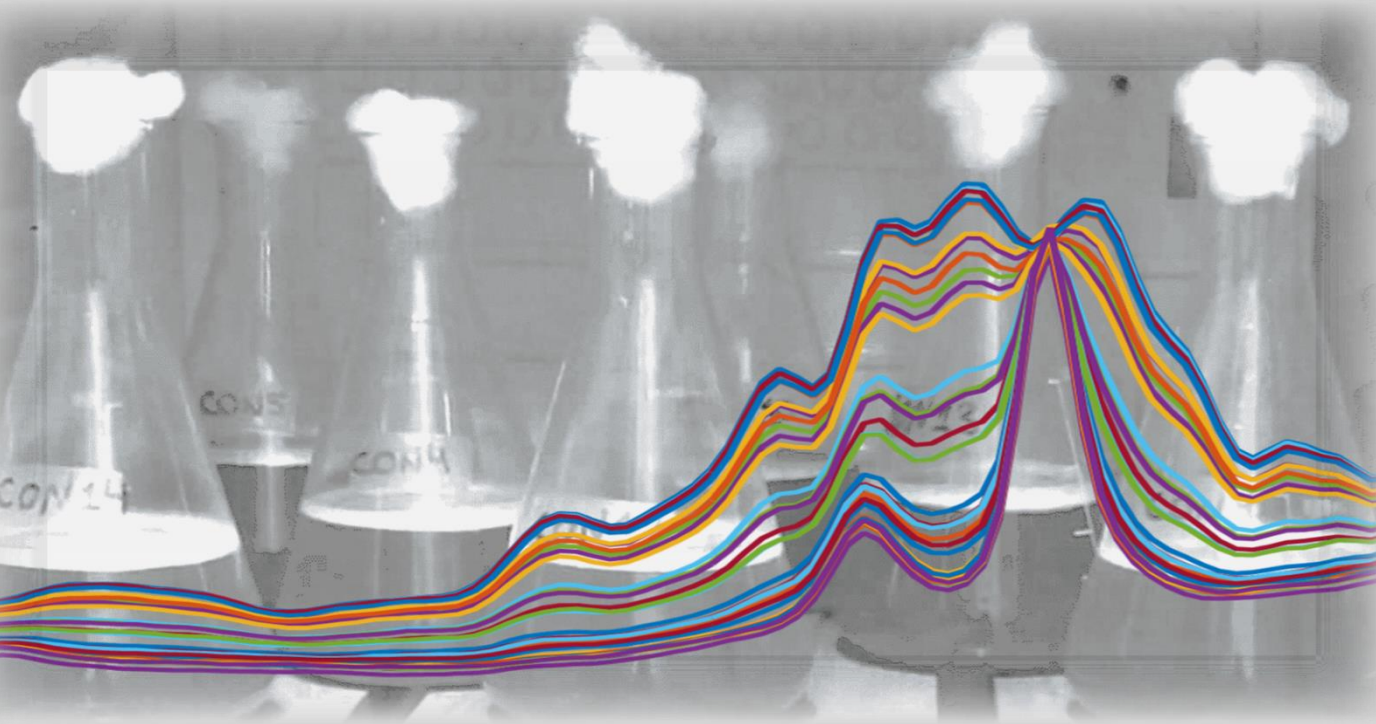
MONITORING WINE FERMENTATION USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES

Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA i VIRGILI

MONITORING WINE FERMENTATION USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES

Julieta Cavaglia Pietro



DOCTORAL THESIS
2021

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA I VIRGILI
Departament de Química Analítica i Química Orgànica
Tarragona, 2021

Julieta Cavaglia Pietro

MONITORING WINE FERMENTATION USING ATR-MIR SPECTROSOCPY AND CHEMOMETRIC TECHNIQUES

Supervised by
Dr. Montserrat Mestres Solé and Dr. Ricard Boqué Martí

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA i VIRGILI
Dept. de Química Analítica
i Química Orgànica

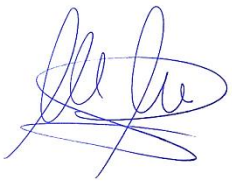
Campus Sescelades, Edifici N4
C/ Marcel·lí Domingo, 1
43007 Tarragona
Tel. +34 977 559 769
Fax +34 977 558 446
www.quimica.urv.cat/qaqo

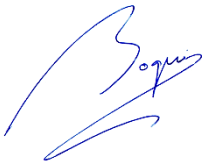
Dr. Montserrat Mestres Solé, Aggregate Professor of Analytical Chemistry, and Dr. Ricard Boqué Martí, Associate Professor of Analytical Chemistry at the Department of Analytical Chemistry and Organic Chemistry of Universitat Rovira i Virgili,

CERTIFY,

that the Doctoral Thesis entitled "**Monitoring wine fermentation using ATR-MIR spectroscopy and chemometric techniques**", submitted by Julieta Cavaglia Pietro to receive the degree of Doctor with International Mention by Universitat Rovira i Virgili, has been carried out under our supervision, in the Department of Analytical Chemistry and Organic Chemistry of this University, and all the results presented in this thesis were obtained in experiments conducted by the above mentioned student.

Tarragona, 28th June 2021,





Dr. Montserrat Mestres Solé

Dr. Ricard Boqué Martí

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Generalitat de Catalunya
Departament d'Economia i Coneixement
Secretaria d'Universitats i Recerca

Unió Europea
Fons social europeu
L'FSE inverteix en el teu futur

The present PhD dissertation has been possible with the support of the *Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya*, the European Union (UE) and the European Social Fund (ESF) (2018 FI_B 00844). The Spanish Ministry of Science and Technology (Project AGL2015-70106-R, AEI/FEDER, UE) is also acknowledged for the financial support given.

AGENCIA ESTATAL DE INVESTIGACIÓN

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Unión Europea
Fondo Europeo de Desarrollo Regional

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

This work is part of the research developed at
the Instrumental Sensometry (iSens) Group
of Universitat Rovira i Virgili.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

A mis abuelas,

"Insanity is doing the same thing over and over again, but expecting different results"
Rita Mae Brown, Sudden Death (1983).

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Index

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chapter 1

Chapter 2

Chapter 1

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Introduction

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

1.1. Wine fermentation and quality control

1.1.1. Wine: a beverage with history

Spain is the third country in the world in terms of wine production after Italy and France, representing the 12.9% of the world production with 33.5 million hectolitres (mhl) in 2019[1]. Despite the fact that the consumption of alcoholic beverages in Spain has undergone a small decline in the last decades, wine, beer and spirits are still considered one of the most important traits of Spanish culture and tradition, as almost 70% of these beverages are consumed in catering establishments and in social environments[2,3].

In Spain, there exist records of wine consumption and grapevine cultivation dating from the third millennium BC[4]. At that time, vine cuttings were imported into the Iberian Peninsula by the Phoenicians, who brought knowledge of viticulture and winemaking to Southern Spain[5].

1.1.2. Grapevine cultivars

The scientific name of grapevine, which is the type of climbing plant on which grapes grow, is *Vitis vinifera* L. and it belongs to the *Vitaceae* plants family. The genus *Vitis* is considered the most important genus in agronomic sciences, being the species *Vitis vinifera* the most widely cultivated and the only species used in the global wine industry[6].

Vitis vinifera L. includes two subspecies: *V. vinifera* L. spp *vinifera* and *V. vinifera* L. ssp *sylvestris*. The latter is the wild form (wild-type), and it is considered the ancestor of the V. *vinifera* spp *vinifera* subspecies, which is the cultivated form[7]. This classification could be argued though, because the morphological differences between these subspecies are most likely the result of human domestication, rather than geographical isolation[6]. Cultivated grapevines with similar, uniform and stable vegetative and reproductive characteristics, which when propagated by appropriate means retain those characteristics, are called "varieties" by growers and "cultivars" by botanists[8]. Nowadays, thousands of *Vitis vinifera* cultivars exist, which have been generated by vegetative propagation and by crosses. In contrast, wild-type cultivars are rare[9]. The reason for this is that domestication of grapes involved several changes in their biology and morphology that ensured greater sugar contents for better fermentation and more regular wine production[6,10].

The global market for wine production is only dominated by few cultivars and they are classified according to their final production: wine grapes, table grapes and raisins[6]. Some

of the most famous cultivars in Spain are *Tempranillo*, *Garnacha* and *Merlot* for red wines, and *Verdejo*, *Moscatel* and *Macabeo* for white wines[11]. Catalonia is particularly famous for its wine tradition, especially for the production of *cava* (a type of sparkling wine made with Catalan grape varieties: *Macabeu*, *Parellada* and *Xarel·lo*) and for the production of wines under 12 Catalan designations of origin (DOs), which are the highest category of Spanish wine regulations. Products within a DO are expected to be of superior quality and to have specific characteristics of geographical region, as well as be derived from raw materials originating within the region. The most internationally recognized Catalan DOs are: Priorat (located in the province of Tarragona), Penedès, Terra Alta and Tarragona[12]. The most cultivated grape varieties in Catalonia include: *Garnatxa blanca*, *Macabeu*, *Parellada* and *Chardonnay* for white wines and *Garnatxa negra*, *Ull de llebre*, *Carinyena* and *Cabernet Sauvignon* for red wines[13].

1.1.3. The wine production process

The understanding of the chemical nature of wine grapes is crucial to provide the wine industry with quality grapes, which will lead to the production of quality wines. Grape quality is affected by the maturity, purity, aroma/flavour and phenolic characteristics, harvesting methods, transportation and processing systems[14]. Traditionally, white wine cultivars aim at obtaining an ideal sugar-to-acid ratio together with moderate must pH and clean varietal character, whereas red wine cultivars are focused on obtaining grapes with rich sugar and anthocyanin content, matching a perfect texture with the grape tannins present in the skin and seeds[15].

For centuries, winemaking has been considered a practical art, and producing wine held great mystery. Making an excellent wine is still challenging, but the scientific principles for understanding wine production, since Louis Pasteur findings on the existence of yeasts, have gradually developed to the science of winemaking: *oenology*. This science encompasses several disciplines: chemistry, microbiology, sensory analysis, biochemistry and engineering[16].

The wine production process will differ depending on the type of wine that is being produced, and it involves a series of steps that need to be mastered and thoroughly controlled in order to obtain wines with the desired quality standards[17]. Many types of wines exist, which can be classified as natural wines (those containing only alcohol coming from alcoholic fermentation) and fortified wines (those that contain alcohol from alcoholic fermentation and also added alcohol)[18]. Figure 1 shows a schematic representation of the

steps involved in the production of natural white and red wines, the two most important types of wine.

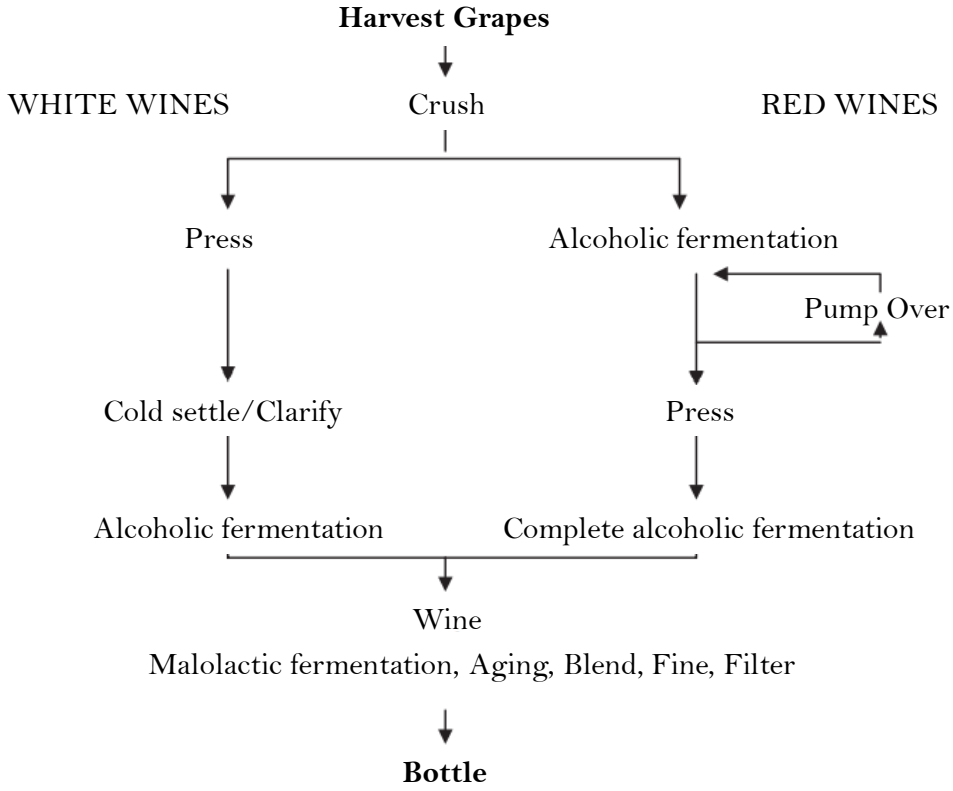


Figure 1. Schematic representation of wine production. Adapted from Bisson, 2004[19].

The main difference between red and white wine production practices is that after harvesting, red grapes are crushed and fermented with grape skins and seeds in order to extract red pigments such as anthocyanins, while white grapes are directly pressed to release the must (grape juice), avoiding any contact with insoluble parts during alcoholic fermentation. In red wine fermentation, the insoluble components tend to form a cap on the surface of the wine once fermentation initiates and this cap must be submerged (pumping over) during fermentation to enhance the extraction of the skin and seeds components as well as inhibiting the growth of aerobes that may be on the exposed surface of the cap. After several pumping over cycles, red wine must is pressed and fermentation is completed. Depending upon the style, the wine is either immediately ready for bottling (preferably under sterile conditions to achieve microbial stability) or it undergoes post-fermentation processing such as aging in wood or stainless steel, a secondary malolactic fermentation, fining to remove undesirable components, or filtration to remove hazes and precipitates[19].

As grapes are the principal raw material in wine, most part of the chemical composition in the final wine will be dominated by the final concentration of the compounds present in the grapes at the moment of harvesting mainly sugars, colour, pH, titrable acidity, organic acids and phenolic compounds[14].

Pulp is the most considerable fraction of the fruit and, in ripe grapes[20]. Sugars (essentially glucose and fructose) are the main constituents of ripe grapes, followed by organic acids (mainly tartaric, malic and citric acids), inorganic anions (phosphoric acid) and cations such as potassium and ammonium, and amino acids. Moreover, the pulp is also characterized by the accumulation of a wide variety of alcohols, aldehydes and esters, which participate in grape aromas[21].

The concentration of fermentable sugars in grape musts at the time of harvest can vary between 180 and 260 $g \cdot L^{-1}$. It is likely that the initial concentration of glucose and fructose selectively influences the type and activity of yeasts present during fermentation. Musts with a lower sugar concentration start to ferment quickly and the sugar ferments to the end, whereas musts with a high sugar content induce osmotic stress in the yeasts and ferment slowly with incomplete sugar consumption[22].

Grape harvesting and alcoholic fermentation are crucial steps in wine production, mainly because wine is especially susceptible to microbiological activity that can alter its quality[23]. In fact, there are microorganisms that are essential for wine (such as yeasts) and microorganisms that can alter the wine positively (some strains of lactic acid bacteria) or negatively (e.g. acetic acid bacteria)[24].

1.1.4. Alcoholic fermentation: the core of wine production

Technically speaking, wine is an ethanolic beverage produced through alcoholic fermentation of grape juice (known as grape must) by the action of yeasts. These microorganisms transform sugars from the juice into ethanol and carbon dioxide (CO_2). *Saccharomyces cerevisiae* is the most important species of yeast involved in the alcoholic fermentation of the grape must[20].

After alcoholic fermentation, the two major constituents in wine are water and ethanol, which represent more than 90% of its total composition[21]. The other wine components are minor compounds that can come directly from the grapes, or by the action of yeasts through their secondary metabolism. During alcoholic fermentation, yeasts are also responsible for the conversion of various chemical intermediates into by-products that contribute to the

characteristic flavour and aroma bouquet of the final wine[25]. To facilitate the complex composition of wine, a common classification of wine compounds is made between volatile compounds (those affecting wine aroma) and non-volatile compounds[26]:

- **Non-volatile compounds**: Sugars, amino acids, organic acids, and phenolic compounds are the most important non-volatile flavour compounds in wine.
- **Volatile compounds**: Hundreds of volatile compounds have been identified and many of them are directly responsible of wine aroma, including alcohols, esters, organic acids, volatile phenols, lactones, acetyls, aromatic ketones, terpenes and fatty acids.

Yeasts are unicellular eukaryotic microorganisms that belong to the Fungi kingdom[27]. Depending on aerobic conditions, yeast can degrade sugars through alcoholic fermentation or through respiration. The primary metabolism, alcoholic fermentation, has been extensively studied due to its industrial applications[20].

Depending on the strain of *S. cerevisiae* used, some characteristics of the wine may be promoted or repressed[28]. On the surface of grapes, both harmful and beneficial species of yeasts are responsible of spontaneous fermentation. However, as *Saccharomyces cerevisiae* species do not naturally impose because they are found in very low concentrations on healthy berries, spontaneous fermentation is usually avoided[29].

Even though some autochthonous yeast species that influence the chemistry of the wine can remain during the early stages of alcoholic fermentation, their growth is limited to the first 2-3 days of fermentation due to their sensitivity to ethanol at concentrations above 5% to 6% (v/v). Under these conditions, strains of *S. cerevisiae* become the dominant yeasts and end the fermentation, as they are more tolerant than other species to ethanol and high concentrations of sugars[22].

The conversion of sugars into ethanol is affected by several factors, including temperature, initial sugar concentration, acidity, nutrient availability, yeast strain and yeast activity.

1.1.5. Problems during alcoholic fermentation

For most oenologists and winemakers, the quality of the wine is determined by its sensory quality and most of the efforts during winemaking are focused on avoiding the production of off-flavours, which can be defined as molecules that add undesirable sensory attributes (both aroma and taste) to wine and can substantially reduce its quality. Some of the

biochemical reactions that produce off-flavours are related to microorganism's strains known as "spoilage microorganisms". Other problems may arise as a result of unusual activities of common microorganisms that may develop at an inopportune moment during the process (bacterial spoilage) or under unfavourable conditions (stuck/sluggish fermentations)[30]. An incomplete use of sugars and nutrients during alcoholic fermentation can lead to the appearance of off-flavours in the wine and increases the risk of bacterial spoilage[31]. The need for more resistant strains to unfavourable wine conditions has led to the production of stress-resistant active dried yeasts starter cultures, which improve nutrient uptake and assimilation, and enhance resistance to ethanol, to other inhibitory molecules and to chemical preservatives (e.g. sulphites) or other antimicrobial compounds[19]. Microorganism spoilage usually takes place at the initial or final stages of alcoholic fermentation, and it is highly undesirable[32]. Monitoring wine fermentation to prevent the appearance of these problems is of utmost importance to the oenologist[33].

1.1.5.1. Stuck and sluggish fermentations

The environment for yeast and bacteria during alcoholic fermentation is very stressful. Yeasts have mechanisms to respond to the evolving environment, but sluggish or stuck fermentations can occur if the yeast cannot adapt to the medium in time or its growth its limited[34].

The main factor that causes stuck and sluggish fermentations is a nutritional limitation, followed by a high concentration of inhibitory substances such as ethanol. Under these conditions, the fermentation performance is inefficient and can be aggravated by the appearance of adverse environmental factors, such as temperature fluctuations that strongly affect the utilisation of nutrients[35]. There are four types of sluggish fermentations, depending on when fermentation begins to slow down: slow initiation becoming to normal, normal initiation becoming sluggish, sluggish throughout the entire process and an abrupt arrest (stuck) at the end of the fermentation[36].

Nitrogen is the most common nutrient that limits wine fermentations; its deficiency can lead to a premature cessation of fermentation or slow down the process, as well as to the production of off-flavours such as hydrogen sulfide. In addition, nutritional limitations such as nitrogen deficiency can prevent yeast population from attaining a typical level of biomass, which is around 10^8 CFU·mL^{-1}. This results in a sluggish fermentation because there are fewer cells fermenting in the medium[37].

Apart from the nitrogen amounts its source must also be accounted. Thus, the preferred source is ammonium but, when it is depleted, yeasts can utilize amino acids (except proline) as nitrogen source. A correct metabolism of nitrogen-containing compounds by yeasts is important, as it produces some end products of sensory importance for wine quality[38].

To avoid stuck or sluggish fermentations, the oenologist can rely on several preventative measures such as ensuring a correct yeast nutrition by the addition of nutrients, using strains with good fermentation potential and monitoring the temperature throughout the fermentation[36]. For this reason, nitrogen supplementation is a common practice especially when using deficient musts, in which the nitrogen concentration is between 120 and 140 mg $N \cdot L^{-1}$ and the complete fermentation of 200 $g \cdot L^{-1}$ of sugar is not guaranteed[39]. However, especial attention must be given to the winemaking strategies and the yeast strains used, as over-supplementation can modify the production of biomass, glycerol, acetic acid and volatile aroma compounds, thus decreasing the quality of the final wine[40].

1.1.5.2. Bacterial spoilage

Part of the microflora present on the surface of the berries is considered to have a significant effect in the early stages of wine production and can have negative effects on the wine as can be seen in Table 1[29]. Preventive measures during alcoholic fermentation to avoid microbiological alterations include: the use of antiseptics (mainly sulphites), the use of selected yeast strains and the thorough disinfection of wine receptacles, equipment and cellar facilities[41].

Table 1. Microflora found on the surface of grapes and their possible negative effect during wine alcoholic fermentation.

Microorganisms	**Negative effect**
Yeasts	Autolysis; spoilage
Lactic acid bacteria	Malolactic fermentation; spoilage
Acetic acid bacteria	Spoilage; stuck fermentation
Fungi	Botrytised wines; mouldy; corky taints
Bacillus species	Spoilage
Streptomyces species	Earthy; musty taints
Bacterial viruses	Destruction of bacteria conducting malolactic fermentation

The most common effects of microbial spoilage are film formation in bulk wines, cloudiness, sediment formation, gas production in bottled wines and formation of off-flavour compounds. The latter effect is the most difficult to detect in time, because some microbial metabolites contribute to wine flavour even when their concentration is below their sensory threshold and can be in the order of nanograms[42].

1.1.5.2.1. Lactic acid bacteria spoilage

Lactic acid bacteria (LAB) are anaerobic and aerotolerant Gram+ bacteria that are found naturally on the surface of healthy grapes. According to their metabolic activity, LAB are divided into three groups: homofermentative, facultatively heterofermentative and heterofermentative[43]. Several LAB species that have been isolated from the grape surface and the must have been associated with wine spoilage. They belong mainly to one of these strains: *Lactobacillus*, *Leuconostoc*, *Oenococcus* and *Pediococcus*[23]. Although lactic acid bacteria can be beneficial for some wines, when acting under uncontrolled conditions these can produce some negative effects such as: excessive deacidifcation, increase of volatile acidity due to acetic acid production, appearance of sourness or synthesis of biogenic amines[44].

1.1.5.2.1.1. Malolactic fermentation

Malolactic fermentation (MLF) consists on the transformation of L-malic acid (a dicarboxylic acid) into L-lactic acid (a monocarboxylic acid) and CO_2 (Figure 2). This transformation can reduce the titratable acidity of the wine by 0.01 to 0.03 g tartaric acid equivalents·L^{-1}, and pH can also be increased by 0.3 pH units. This results in a deacidifcation of the wine, which has a major impact on the quality, flavour and aroma of the wine[45].

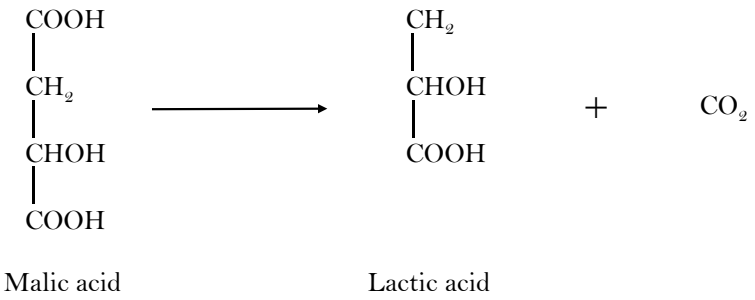


Figure 2. Conversion of malic acid to lactic and CO_2 during malolactic fermentation. Malic acid has two acidic groups that can release protons, while lactic acid only contains one proton that can be released.

Secondary compounds are also generated through this fermentation process, bringing important changes in both the wine quality and composition and providing microbiological stability to the final wine[46].

The most common bacteria strain used for oenological practices is *Oenococcus oeni*. This strain can easily impose when sugars are almost depleted, and it tolerates high amounts of alcohol. They proliferate better if there are low levels of free SO_2 in the wine. In addition, some of the products from the secondary metabolism of LAB soften the wine and add aromatic notes. Ethanol, acidic pH, phenolic compounds and sulphur dioxide can induce stress on LAB[47].

Traditionally, malolactic fermentation is desirable in red wines and high acidity white wines because MLF increases wine pH and produces some aromatic compounds that improve the organoleptic profile of the final wine[48]. Spontaneous MLF is generally discouraged, and an starter culture of LAB is added to avoid the growth of indigenous LAB that may produce negative compounds in the final wine[49].

There are three possible inoculation stages for MLF:

- **Pre-fermentation inoculation**: In this stage, LAB can decrease nutrient availability for yeasts, resulting in stuck/sluggish fermentations. In addition, LAB can produce compounds that are inhibitory to yeast growth and fermentation.
- **Simultaneous with Yeast Inoculation**: Co-inoculation allows LAB to acclimate more easily to the must media, as ethanol concentration is low and the availability of nutrients is high at this stage. This increases the adaptation of LAB to high ethanol levels in wine. However, inoculating at this stage can produce an increase in the production of acetic acid, and a decrease in the viable populations of both bacteria and yeast can be observed. This is due to the competitiveness for nutrients between the microorganisms. Some commercial yeast strains are more sensitive than others to fermentation arrest upon inoculation of LAB.
- **Mid-fermentation inoculation**: Inoculation during alcoholic fermentation can be dangerous for the viability of LAB, as nutrients are limited due to yeast activity. Bacterial biomass may also deplete the fermentation of needed survival factors. At this point, the advantage is that the content of ethanol is still low and not inhibitory to MLF and most SO_2 is combined.
- **Post-fermentation inoculation**: This strategy prevents any inhibitory effect of LAB against the yeast and allows better temperature control. However, ethanol

concentration may be too high and nutrients are exhausted, so that LAB acclimation in these conditions is difficult.

The most common decision is to inoculate LAB at the end of the alcoholic fermentation, to avoid an excessive development of LAB that can yield high levels of acetic acid. In addition, co-inoculation and inoculation during alcoholic fermentation tends to prolong MLF, suggesting that inhibition to LAB activity exists. Generally, the yeast-bacteria interaction must be closely studied, as certain strains of these microorganisms can either inhibit or stimulate alcoholic fermentation or MLF[50].

1.1.5.2.2. Acetic acid bacteria spoilage

Acetic acid bacteria (AAB) are alfa-proteobacteria and they belong to the family of acetobacteria. They are commonly known as the vinegar bacteria and they can be naturally found in substrates rich in sugar and ethanol, such as fruits, flowers, foodstuffs and alcoholic beverages[51]. AAB are gram-negative and aerobic bacteria, and some of them can survive in wine, where they can rapidly oxidise glucose and ethanol as sources of energy[52].

Depending on the sugar concentration and the pH in the must, AAB can oxidise sugars or ethanol to produce acetic acid as the end-product through a process called acetification[53]. In healthy grapes, where the global AAB population is around 10^2 to 10^3 CFU·mL^{-1}, the most common AAB strain present on the surface of the grape is *Gluconobacter oxydans*. In contrast, AAB population in damaged grapes can reach up to 10^6 CFU·mL^{-1}, with *Acetobacter aeri* being the most present strain. When alcoholic fermentation begins, the medium conditions are unfavourable for most AAB strains. This is why most AAB found on the surface of grapes disappear as soon as alcoholic fermentation begins. However, *Acetobacter aceti* and *Acetobacter pasteurianus*, which are strains more tolerant to ethanol, can be found in fermenting musts and may survive due to a poor management (e.g. exposure of wine to air during pumping and transfer procedures, or contamination in wooden barrels) throughout alcoholic fermentation, having a negative influence on the final wine[23]. Growth of acetic acid bacteria has also been observed in bottled wine, mainly because the wine was bottled without sterile filtration and/or the concentration of sulphur dioxide was too low prior to bottling[54].

1.1.6. Wine Quality Control

Wine quality can be defined from two perspectives: sensory or chemical, and it is difficult to measure. From a sensory perspective, wine quality assessment has traditionally relied on the sensory evaluation by wine experts (panellists) and winemakers, as they have the training and experience to detect wine defects and are capable of performing perceptual evaluations[55]. The problem of human assessment is that it is extremely subjective and depends on preferences of panellists and their individual taste and aroma limits of perception among other factors. There are several methodologies to alleviate subjectivity in evaluating wine quality which focus on the training of panellists both on recognition of descriptors or sensory attributes and in their evaluation[56]. The tasting room, environmental conditions and equipment are other parameters to be considered and can be designed according to specific guidelines for sensory analysis as, for example, ISO indications which become mandatory in the case of accreditation for quality purposes[57]. An analytical sensory panel is generally formed by 8 to 20 individuals, who must be trained to perform sensory tasks objectively and consistently[58].

From the chemical perspective, among hundreds compounds found in grapes and wines, there are some molecules that are considered as reference parameters for quality control. This is because their presence and/or concentration is related to a specific metabolic pathway or chemical behaviour of the sample. Thus, the continuous measurement of quality control parameters through the production process, especially during alcoholic fermentation, ensures a successful wine production. Chemical analyses of grapes and wine are essential to ensure that the different steps of winemaking take place under the desired conditions. However, the time between grape sampling and its arrival to the laboratory should be short enough to ensure that, if necessary, some adjustments can be made to the process. Monitoring the progress of fermentation ensures that the wine is under control and complies with legal specifications.

There are a variety of analytical methods to analyse the different reference parameters that are used to assess wine quality. Methods range from classical chemical analysis to modern instrumental techniques. The International Organisation of Vine and Wine (OIV) is the organism responsible for the publication of official methods for the analysis of grapes, musts and wines, which are detailed in the Compendium of International Methods of Analysis of Wines and Musts. The Compendium aims at standardizing the methods of analysis at an international level, and includes the only accepted methods for specific parameters, as well

as benchmark methods and recommended alternative methods in cases of disputes and for calibration purposes[59]. It also includes methods for inspection and regulatory purposes. Auxiliary methods of recent implementation, in which the method performance has yet to be established, are also be described[60].

It is not surprising that, due to the complexity of wine, the number and variety of analytical methods found in OIV Compendium is large. Methods are divided into physical (e.g. density, specific gravity, total dry matter) and chemical analysis (e.g. sugars, alcohols, acids, gases and non-organic compounds). Methods range from classical techniques (e.g. colorimetric assays, distillation, etc.) to more modern methods such as the determination of mineral elements by inductively coupled plasma atomic emission spectroscopy (ICP-AES)[59]. In practice, very few wine cellars cover all these analyses at all the steps of wine production, and a selection of key parameters that require a continuous checking must be made which usually are sugars, acids and nitrogen.

1.1.6.1. Traditional fermentation monitoring

For the purpose of assessing microbiological stability, alcoholic fermentation is controlled by means of analytical tests of the following parameters[41]:

- **Reducing sugars:** they serve as an indicator of the consumption rate of yeasts. At the end of the alcoholic fermentation, their presence can cause microbiological instability by supporting the growth of yeast and bacteria. The consumption of sugars is usually monitored by densimetry.
- **Alcohol content** Wines typically have a 10-14% w/w ethanol. Ethanol serves as a preservative of wine spoilage, as some bacteria are not resistant to high ethanol level. A low alcohol content can increase the risk of bacterial spoilage.
- **Total, free and molecular sulphur dioxide:** The use of SO_2 is critical to prevent bacterial spoilage in wine. However, recent health concerns against the use of SO_2 and its legal maximum limits in wine are leading towards the interest of minimizing the amount of SO2 in wine[33,61].
- **pH:** it normally ranges from 2.9 to 3.9 and it can be measured using a pH-meter. During alcoholic fermentation, there is a decrease in pH due to yeast activity. An increase in pH is an indication of alterations in acidity due to malolactic fermentation. Instead, a decrease in pH indicates advanced stages of acetic acid spoilage. Tartaric acid precipitation can also cause a change in the pH.

- **Volatile acidity:** an increase after alcoholic fermentation can indicate bacterial spoilage.
- **Total acidity:** it decreases notably in case of malolactic fermentation. It increases in the case of acetic acid or lactic acid spoilage.
- **Malic Acid and lactic acid:** the simultaneous disappearance of malic acid and formation of lactic acid is an indicator of malolactic fermentation. At the beginning of alcoholic fermentation, there is always a decrease in the malic acid level due to maloalcoholic fermentation by yeast. This decrease does not usually exceed 25% of the initial malic acid concentration.
- **Titratable acidity:** it is a measure of the total acid content (except CO_2) in wine. It can be determined by classical titration with an indicator or with a photometric analyser. It is usually expressed as $g \cdot L^{-1}$ equivalent of tartaric acid.

Even though chemistry knowledge for the viticultural and oenological practices has made a remarkable progress in the last decades, small cellars might not have the economic capacity or infrastructure to analyse all these parameters in their facilities. Some cellars may send samples to external laboratories, but others may only rely on basic parameters to control the fermentation process (mainly reducing sugars and pH) and trust the expertise of the oenologist. In addition, many important sensory compounds cannot be detected or measured quickly, efficiently and inexpensively during the winemaking process.

1.2. Process Analytical Technologies

1.2.1. The idea behind PAT

In the past, biopharmaceutical production was focused on the analysis of a target product, and the process was merely limited to the definition of a series of operating conditions that were required to achieve consistent product quality. The Quality by Design (QbD) principle, in contrast, states that the quality of products can and should be incorporated by process design, building quality in process design and improving process understanding, instead of post-production testing for quality[62]. This idea has been an emerging topic not only in the biopharmaceutical industry but also in the food industry, where the design of the process to deliver products with significant attributes of safety, quality and/or efficiency is reaching special importance[63].

To ensure process performance, a robust control strategy must be implemented. Process Analytical Technologies (PAT) facilitate the implementation of QbD, ensuring process

control by an ongoing monitoring of process performance[64]. The PAT concept was first introduced in the USA for the pharmaceutical industry in 2004, by the American Food and Drug Administration (FDA). PAT is defined as "*a system for designing, analysing and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality*"[65].

The application of PAT implies the transition from a laboratory-based quality control to a process-based quality control, moving directly "*into the field*" to the site of production[66]. In-process measurements allow the permanent monitoring of the product or process conditions, providing lower risks of releasing a poor quality product and a high degree of process understanding and efficiency through continuous learning and improvement[66,67]. This idea results in the change from post-problem or feed-back process control to during-problem control, where process adjustments can be made during manufacturing to compensate, for example, unwanted variations in raw materials. PAT emphasizes the anticipation of problems, rather than dealing with them *post hoc*[68].

The most popular analytical techniques recommended for PAT are Near infrared (NIR) and mid infrared (MIR) spectroscopy, followed by electrochemistry and chromatographic techniques[69]. It comes as no surprise that spectroscopic methods are the most suitable for PAT; their speed, flexibility, simplicity, non-destruction of the sample and portability make them the perfect choice for process measurements and non-invasive analysis[70].

1.2.2. Types of PAT measurements

During product processing, different types of analysers can be used that provide information related to biological, physical, and chemical attributes of the process[65]. Depending on the physical characteristics of the analyser and how the measurement of the sample is obtained, we find:

- *At-line* measurements: the analyser is located in the production area, near the process stream. Benchtop instruments and portable instruments meet this condition. They usually require an operator, and the sample is removed and isolated from the process stream, and finally analysed. These measurements may require fast pre-treatments or no pre-treatment at all[71].
- *In-line* measurements: they require probes or sensors that are placed directly in the process stream. They can be assembled in the production line, and the sample is

not removed from the process stream, allowing to analyse the sample in the most real possible conditions[72].

- *On-line* measurements: this strategy is very similar to the *in-line*, the only difference is that the sample must be diverted from the process stream and may be returned to it after the measurement[73].

A schematic outline of the different ways to implement process control is shown in Figure 3. It is worth mentioning that both *in-line* and *on-line* measurements are usually fully automated. In addition, although the human error factor is removed, new errors related to the automation of the process may appear, and thus special attention must be paid to the development of these fully automated processes, in order to avoid electronic and computer errors[74].

Off-line measurements, in turn, imply that a sample from the process stream must be taken to an external laboratory for its analysis. This strategy is often used when the number of batches in a process is low and the analysis is difficult to automate. Analytical results from *off-line* measurements do not provide *real-time* results, as the process status information is obtained with an implied time delay and; therefore, these measurements cannot be considered as a PAT tool[75].

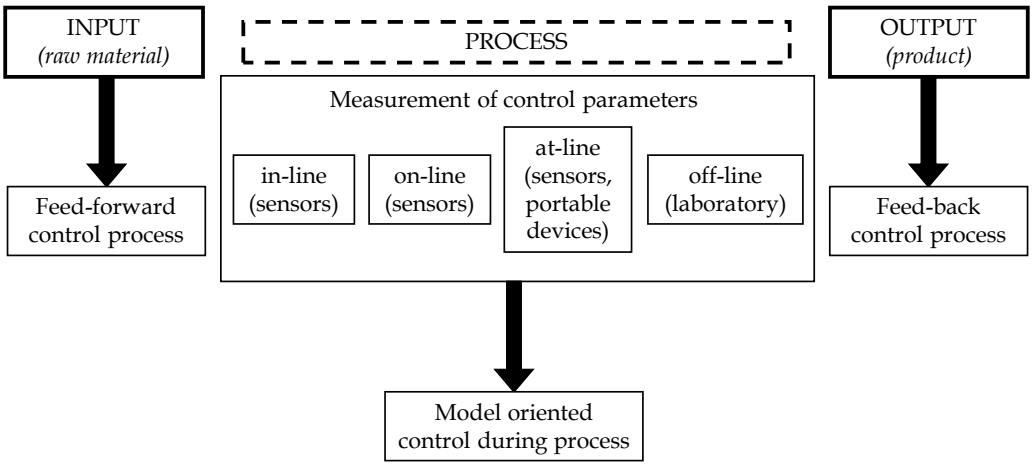


Figure 3. Outline of the different approaches that can be implemented for process control, specifying those possible measurements during processing.

Great efforts have been made to develop analytical instruments that bring the analysis closer to the process, making a switch from the laboratory to the location where the sample is. To this aim, handheld spectrometers have emerged as improved measurement tools for the qualitative and quantitative chemical quality control, not only in the pharmaceutical sector but also in many different industrial branches[76]. Nowadays, the tremendous advances in sensing technologies have even made it possible to develop smartphone-based optical biosensors[77].

1.2.3. PAT in the food and beverage industry

With the pharmaceutical industry as a starting point, PAT has emerged as a platform for future good manufacturing practice (GMP) to other industry branches, introducing new concepts to different sectors regarding the monitoring and control of manufacturing processes. One of these sectors is the food and beverage industry, where the introduction of PAT has led to an improvement in process control, final product quality evaluation, productivity and optimization of raw materials[78].

Traditionally, in the food industry the quality of a product is evaluated through periodic chemical and microbiological analysis, following the HACCP (Hazard Analysis and Critical Control Point) principles, which were developed by the Codex Alimentarius Commission, in combination with the ISO 22000, to meet the needs of all organizations participating in the food supply chain[79]. Once the hazards of a product have been identified, control measures for each hazard must be defined, as well as the frequency of their application, the sampling procedures and the criteria for product acceptance[80].

In line with PAT principles, it is now possible for food manufacturers to deliver a product that does not necessarily require post-process testing, through continuous *at-line* testing of the products during production. Nowadays, the use of fast analytical techniques has an increasing role in the food industry, from the most classical analysis of quality control parameters in foodstuffs throughout their processing[81,82]. to more unusual analyses, such as the detection of fraud in value-added products[83]. One major advantage of using PAT tools, and especially spectroscopy-based process analysers, is that quality parameters or even factors that have an influence in the final product can be directly predicted by establishing a statistical relationship between the measured signals and reference analysis[67].

When defining the most effective quality control strategy to be implemented in the agro-food sector under the PAT guidelines, it is important to thoroughly study the composition

and main characteristics of the raw materials concerned. One of the major challenges for the application of PAT in the food industry using vibrational spectroscopic techniques is that food samples are usually heterogeneous mixtures of complex biomolecules in different physical matrices (e.g. aqueous, solid, gels, etc.) and that these matrices usually evolve during processing (e.g. fruits ripeness, fermentation, ageing, etc.)[67]. First, the key quality parameters, which are the main chemical components or physical properties that define the quality of raw materials, must be determined. The measurement of these parameters, both in the final product and during processing, allows to assess the state of the process and the characterization of the raw materials and food products[84]. Then, all the factors influencing the process should be studied and included in a detailed experimental design whenever possible, as they can critically influence the conditions of the food matrix and, consequently, the analytical response[85].

The application of PAT has been studied in a wide variety of food and beverage products, such as meat[86], dairy[87,88], olive oil[89], bakery products[90] and fruits[91,92]. The use of spectroscopic methods in the wine industry is considerable, to the point that the OIV has published a series of guidelines for the use of infrared analysers in oenology, which describe the quality control procedures to obtain reliable results[93]. Hence, the application of PAT methodologies in the wine industry could be advantageous to obtain wines with the desired characteristics and reduce the risk of encountering problems during the different stages of wine production.

1.2.4. PAT for bioprocess monitoring

Processes involving the action of microorganisms (bioprocesses) are complex, time dependent, and highly nonlinear. These characteristics make these processes very difficult to systematise[94]. Of all food processing technologies, fermentation is one of the oldest and most well-established methods for the production of alcoholic beverages and basic food commodities such as bread and dairy products[95]. *Real-time* monitoring and control of fermentation processes appears to be quite useful for the food and beverage industries, in order to increase productivity, efficiency, and reproducibility[96].

Bioprocess control and monitoring involves the application of sensor or portable technologies for some traditional critical process parameters, including temperature, pH, and dissolved oxygen[97]. However, the process information that can be continuously obtained from these sensors is nowadays deficient, as new parameters, such as matrix composition, biomass concentration or product concentration, that yield valuable

information of the process and have an influence on the critical attributes, should also be controlled during processing. Consequently, research on bioprocess control has been expanding beyond the more traditional control parameters to include more sophisticated process information[97].

Traditional methods usually rely on conventional techniques, such as titration, chromatography, ultraviolet-visible spectroscopy and electrophoresis, among others. As mentioned in the previous section, these analytical techniques are not useful for continuous process monitoring because they are *off-line*, that is, they require that a sample is withdrawn from the bioreactor and, in some cases, steps of extraction or laborious sample pretreatment are needed, thus increasing the time of analysis[98].

There is a growing interest in the implementation of fast and affordable analytical methods in the food and beverage industries, as new instruments allow gaining process knowledge through the determination of compounds that could only be previously monitored using traditional techniques[99]. In this regard, Near-Infrared (NIR) and Mid-Infrared (MIR) spectroscopy have gained popularity in bioprocess monitoring during the last decades[72,96]. These techniques offer the possibility to obtain a significant amount of information about the process, which could be considered as "fingerprints" of the different stages of the process[72].

1.2.5. NIR and MIR for bioprocess control

IR and MIR spectroscopies are ideal techniques for bioprocess monitoring, as they can simultaneously analyze a wide variety of organic compounds, including substrates, products, metabolites, nutrients and biomass concentrations. Some of the compounds that have been successfully determined with these techniques are: glucose, glutamate, fructose, glutamine, lactate, ammonia, glycerol, acetate and ethanol[100]. These measures have made it possible to implement process control strategies based on substrate depletion (e.g. rate of sugar consumption) or rate of product formation. When monitoring bioprocesses with IR techniques, the most challenging aspects to be considered are the huge matrix differences between the initial and final states of the sample, and the intrinsic variability among different batches of the same process. Both techniques have their advantages and disadvantages, and will be described in the next section. Even though MIR has gained popularity in recent years, its use is less widespread than NIR[70].

The main bioprocesses studied by using NIR and MIR techniques are those related to the detection of substrates and metabolites in yeast and bacterial cultures and in suspended or immobilised cell cultures. It should be noted, however, that most studies have been conducted in synthetic media[101]. Many authors have proposed the use of NIR or MIR to monitor the evolution of organic compounds involved in alcoholic fermentation processes[71,84,102]. The main molecules analysed were glucose and ethanol[103], but spectroscopic methods have proven to be efficient to monitor other important substrates and metabolites such as lactose, galactose, lactic acid and pH[104].

From NIR and MIR measurements, a large amount of data is obtained that requires the use of Multivariate Data Analysis (MVDA) techniques, such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression. These techniques are essential to gain process knowledge as the allow the interpretation and analysis of data[105].

1.3. Spectroscopy

1.3.1. Principles of Spectroscopy

Spectroscopy, one of the most important techniques in analytical chemistry, deals with the production, measurement, and interpretation of spectra arising from the interaction of electromagnetic radiation with matter. Spectroscopic methods provide valuable information on the molecular features of a wide variety of compounds, and it is widely used for both quantitative and qualitative analyses[106]. Electromagnetic radiation is the result of the combination of electric and magnetic fields (Figure 4)[107].

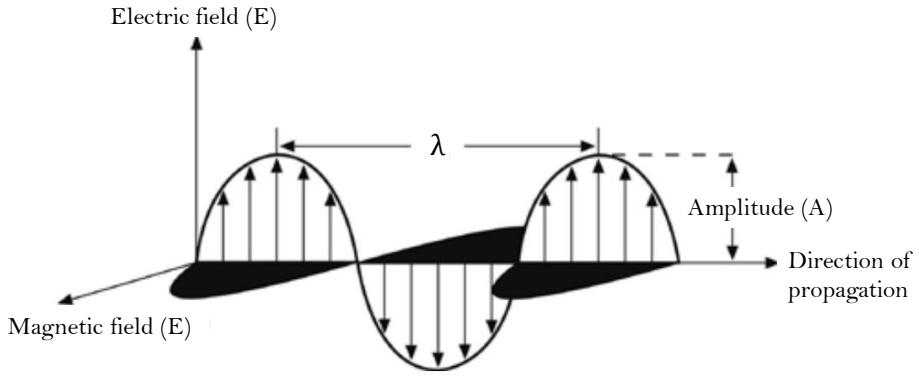


Figure 4. Representation of electromagnetic radiation. The electric and magnetic fields are in phase, perpendicular to each other and to the direction of propagation. Adapted from Penner, 2017[107].

Radiation waves are defined by their wavelength (λ, length of one wave), frequency (v, number of cycles per unit time), and wavenumber ($\bar{\nu}$, number of waves per unit length). These parameters are related to one another by the following expression:

$$\bar{\nu} = \frac{v}{(c/n)} = \frac{1}{\lambda} \qquad \textbf{Equation 1}$$

where c is the speed of light, n the refractive index of the medium where the radiation is passing through and v is the photon frequency. In fact, when interacting with matter, radiation acts as particles of energy (photons), giving rise to the wave-particle duality theory, which states that radiation acts, both at the same time, as a discrete particle and a continuous wave[108]. Photon frequency (v) and photon energy (E_P) are related by:

$$E_P = hv \qquad \textbf{Equation 2}$$

where h is Planck's constant (6.6256 x 10^{-27} erg sec). Photons of specific energy may be absorbed or emitted by a molecule, resulting in a transfer of energy. Depending on the energy of the electromagnetic radiation, transition between different types of energy states are induced at atomic and/or molecular level. The set of different levels of energy radiation constitute what is known as the electromagnetic spectrum (Figure 5), which classifies radiation according to wavelength/frequency values[108].

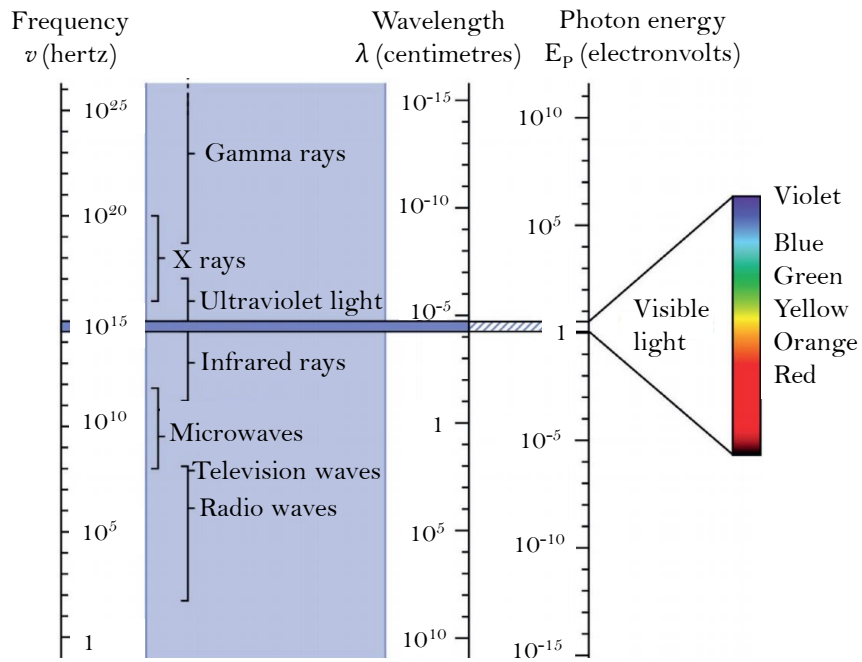


Figure 5. The electromagnetic spectrum. From Odularu, 2020[109].

Different analytical methods exist depending on the region of the electromagnetic spectrum used in the analysis, the type of radiation-matter interaction to be monitored (such as absorption, emission, or fluorescence) and the species to be analysed (such as molecules or atoms)[107].

1.3.2. The infrared region

The infrared region of the electromagnetic spectrum involves radiation waves with longer wavelengths and lower energy than visible light. The energy provided by the photons associated with the infrared region results in transitions between quantized vibrational energy states of molecules, inducing vibrational excitations of bonds or groups within molecules[78]. The identification of the absorption bands in an IR spectrum of specific chemical groups can be used for quantitative and qualitative molecular analysis. Infrared radiation can also cause rotational movements of molecules, but these are generally superimposed on the vibration bands and can only be observed with high-resolution spectrometers[110].

The infrared spectrum of a sample is obtained by passing a beam of infrared radiation through a sample and determining what fraction of the incident radiation is absorbed, transmitted or reflected by the sample at a given energy (represented by wavenumbers or wavelengths that are proportional to the energy as shown in Equations 1 and 2)[111]. Only asymmetric bonds with a dipole moment that changes as a function of time are capable of absorbing IR radiation[112]. The resulting signal represents the molecular infrared absorption, transmission or reflection of the sample. The most interesting feature of these IR signals is that two samples with different composition do not have the same IR spectrum, which is unique for each specific sample and can be seen as the sample fingerprint in the IR region[113].

1.3.3. The IR instrument

There are two types of IR spectrometers: dispersive and Fourier transform (FT) instruments. Back in the 1940s, dispersive systems were the first IR spectrometers available, and they splitted the individual energy frequencies emitted from the infrared source. This was accomplished using a prism or a grating as dispersive elements. These dispersive systems contain elements similar to those of ultraviolet–visible (UV–Vis) spectrometers, including a radiation source, a monochromator, a sample holder, and a detector connected to an amplifier system to record the spectra. In this system, the detector

measures the amount of energy at each frequency that has passed through the sample and the resulting signal is plotted in the form of a spectrum of intensity vs. frequency[114].

The advances in IR technology lead to the appearance of Fourier Transform infrared spectrometers (FT-IR). This type of system was mechanically simpler as they only have one moving part (the interferometer), replacing in this way the dispersive instruments from the original IR instruments. This change drastically improved the quality of the spectra and decreased the time required to obtain an IR spectrum[115].

1.3.3.1. Michelson interferometer

The FT technique uses a Michelson interferometer (Figure 6), an optical device whose main parts are: the source of infrared light, the beam splitter that splits the incoming IR beam from the source into two beams, and two perpendicular plane mirrors: one moving mirror and a fixed mirror[116].

The two beams generated from the beam splitter are reflected into their respective mirrors: one beam is reflected in the fixed mirror, while the other beam is reflected in the moving mirror. Then, they recombine when they meet back in the beam splitter. Because the beam reflecting on the fixed mirror moves at a fixed path length and the other is constantly changing its path length, the signal coming out of the interferometer is the result of the two beams interfering with each other[117].

When the two light beams have travelled the same distance from the beam splitter (zero path difference) the reflected light beams are in phase and all frequencies will interfere constructively to produce the highest intensity (centerburst). Due to the movement of the moving mirror, the position of the two mirrors modifies the distance travelled by the two light beams creating a continuous modification in the optical path difference between the two beams. This creates interferences in the radiation, and thus different interference patterns (intensities) are obtained. The two beams are only in phase when the optical path difference between the mirrors is an integer (n) multiple of the wavelength (λ). If the mirrors are not in phase, the interference is destructive, leading to a beam of low intensity[116].

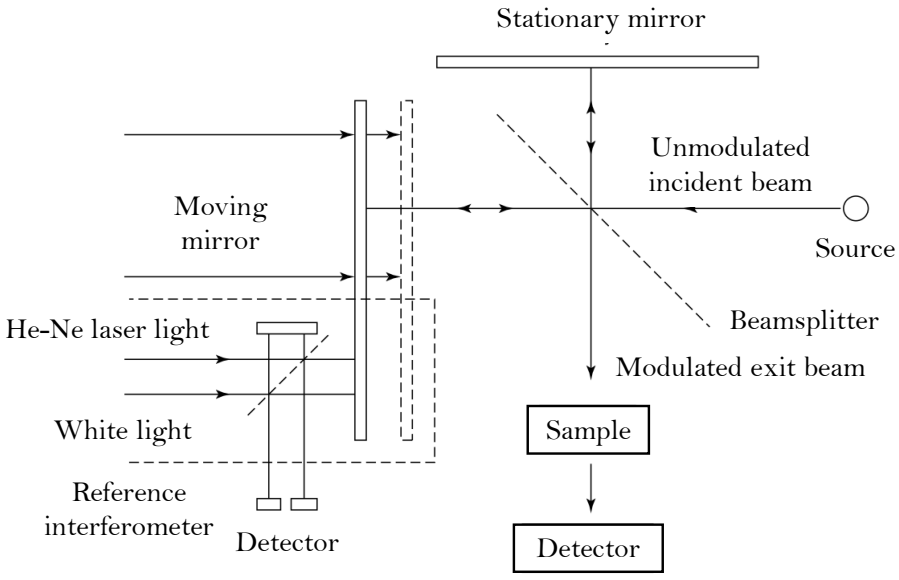


Figure 6. Schematic of a Michelson interferometer. From Stuart, 2005[116].

After the beam has passed through the sample, the resulting signal is called an interferogram (Figure 7). It is a unique type of signal where every data point has information of every infrared frequency coming from the source. The total intensity of the source (in volts) is shown by the centerburst, which does not contain any signal from the sample. The wings of the interferogram contain the information from the sample[106].

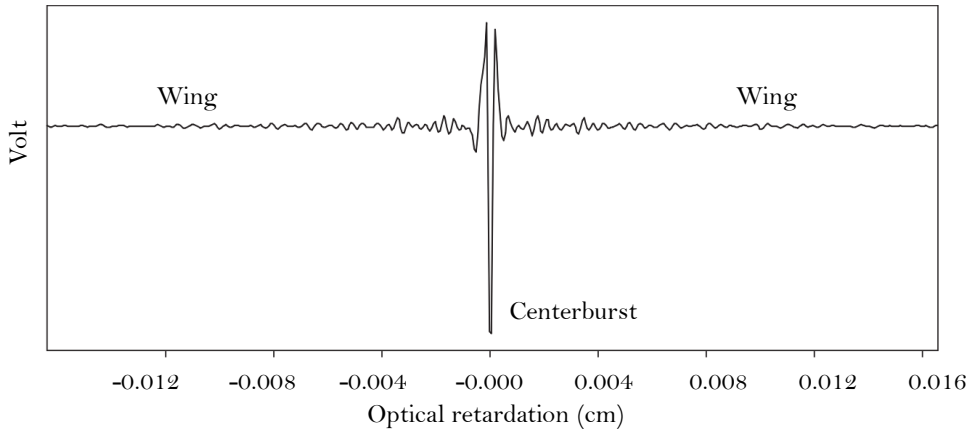


Figure 7. Typical interferogram of a modern FTIR spectrometer. From Sun, 2009[106].

1.3.3.2. Fourier Transform Infrared Spectroscopy

The interferogram resulting from the IR spectrometer cannot be directly interpreted. This signal must be transformed using a well-known mathematical technique called 'Fourier Transformation' in order to get the final spectrum. This calculation is performed directly by the computer using the FT algorithm[96,118]. and consists on breaking down the interferogram into sine waves for each wavelength or wavenumber[106].

The application of Fourier Transformation in IR (FT-IR) has dramatically improved the acquisition of infrared spectra and has given infrared spectroscopy the possibility to[118]:

- Increase the speed of analysis, as all the frequencies are measured simultaneously.
- Perform an internal calibration: these instruments employ a special laser as an internal wavelength calibration standard.
- Increase sensitivity and signal stability: the optical throughput is much higher, resulting in much lower noise levels, and the fast scans enable the coaddition of several scans in order to reduce the random measurement noise to any desired level.
- Reduce the possibility of mechanical breakdown thanks to its mechanical simplicity.

A background spectrum must be measured, usually with air or water (when dealing with samples with high % of water) to determine the relative scale for the absorption intensity. Then, the sample is measured and both measurements (sample and background) are compared to determine the "percent transmittance/absorption". The spectral features obtained in the final spectrum are strictly due to the interaction of the light beam with the sample[119].

For all these reasons, involving simplicity, sensitivity and versatility advantages, this technique is rapidly gaining importance for the analysis of foods and beverages[120].

1.3.4. NIR, MIR and FIR Spectroscopy

Infrared spectroscopy can be subdivided into three categories (Figure 8) according to the level of energy: far-infrared (FIR, <400 cm^{-1}), mid-infrared (MIR, 4000 to 400 cm^{-1}) and near-infrared (NIR, 13000 to 4000 cm^{-1}). Nowadays, most infrared applications employ the NIR region, which consists of overtones and combination bands, but the mid- and far-infrared regions also provide important information about certain materials. Due to the

minor energetic distances between the energy levels in NIR, the resulting spectra are less defined, whereas MIR spectra exhibit a higher absorption capacity and well-defined peaks[96]. MIR has a shorter path length compared to NIR and this reduces light scattering, which is commonly experienced when using NIR. On the other hand, this means that MIR does not penetrate as much into the material as NIR does[70]. Also, MIR has a higher resolution, so it enables detecting and quantifying components in aqueous solutions at lower concentrations than with NIR[96].

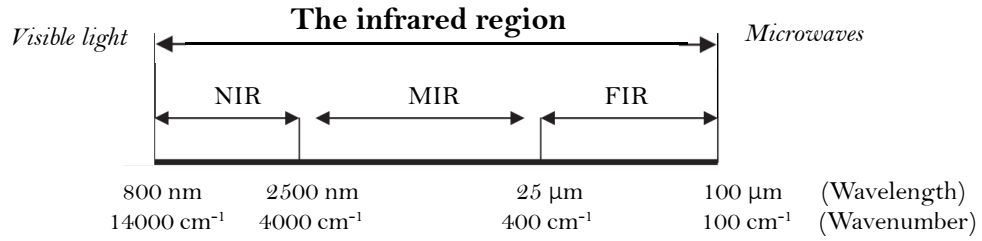


Figure 8. Types of radiation in the infrared region of the electromagnetic spectrum. Adapted from Sun, 2009[106].

FIR spectroscopy is more limited than NIR and MIR for quantitative and qualitative analysis of food-related materials, but it does provide information regarding the vibrations of molecules containing heavy molecules, such as inorganic and organometallic substances[116]. FIR can be used to detect intramolecular stretching, bending and torsional modes involving heavy atoms[121]. Hence, this low-energetic region is not as useful as the other regions for the analysis of the organic constituents in food systems[122].

1.3.4.1. Near-Infrared Spectroscopy

In NIR spectroscopy, the spectrum generated is the result of combinations and overtones of fundamental vibrations of molecules containing hydrogen atoms covalently bonded to carbon, oxygen or nitrogen: -CH, -OH, -NH groups. These groups are largely found in organic molecules such as proteins, carbohydrates and fats, which are the basis of food and beverage products. This characteristic makes NIR a very useful technique for the analysis of foods and beverages[123].

NIR spectroscopy is nowadays an established method for the physicochemical analysis of a wide variety of food products. It has been used for the determination of the concentration of fat, protein or sugars in dairy products, meat and cereals, the analysis of quality parameters such as moisture and tenderness, and the determination of the degree of

ripeness in some fruits[123]. In the beverages industry, it has been used for the analysis of beer, wine and spirits[106].

1.3.4.2. Mid-Infrared Spectroscopy

sIn MIR spectroscopy signals arise from fundamental vibrations and overlap less than in NIR spectroscopy. In MIR spectra, the signal is characterized by absorption due to molecular vibrational and rotational motions. A representation of these vibrational motions inducing this type of energy absorption in the MIR region is shown in Figure 9[114].

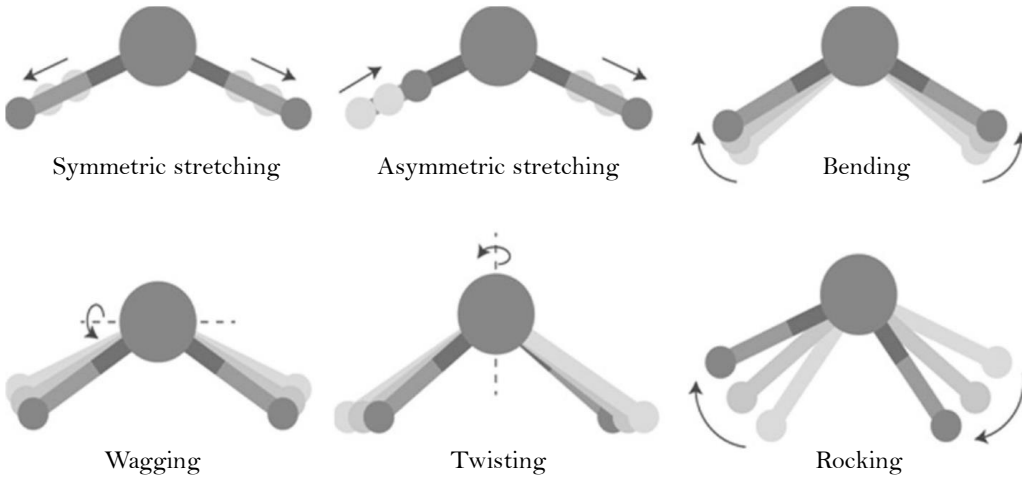


Figure 9. Vibrational modes of the water molecule Stretching vibrations refer to those vibrations in which bond length changes, whereas bending vibrations refer to those vibrations in which bond angle changes. Adapted from Rodriguez-Saona *et al.*, 2017[114].

An interesting feature of MIR spectra is that they contain a region which is specific for each sample. This region is called the fingerprinting region, and it ranges from 1500 to 1000 cm^{-1} and is widely used for the identification of molecular structures[124].

Unlike NIR, MIR bands can be assigned to functional groups, and the intensity of the bands is related to the nature and polarity of the bonds present in the molecule. For example, the C-O bond (highly polarized), strongly absorbs in the MIR region while the C-C bond absorbance in the MIR region is much weaker. Also, stronger bonds vibrate at higher frequencies than weaker bonds and double or triple bonds are stronger than simple bonds (this is determined by the strength constant k of each bond)[116].

These characteristics make it possible to analyse a wide variety of complex samples, including those from the food industry. One of the major drawbacks when analysing food samples with MIR, however, is the strong absorption of water that can overlap or mask

other important peaks. Another drawback is that organic molecules have similar chemical structures, which makes it especially challenging to distinguish specific molecules[113]. This is mainly due to the presence of C-C, C-H and C-O in most of these molecule backbone structure. In food applications, MIR has been used for the determination of fat in dairy products, water, sugars and ethanol in alcoholic beverages, and authentication and quality control purposes[110].

1.3.5. IR analysis methods

New developments in sampling techniques, thanks to innovation in optical probes as well as in electronic/computational techniques, have made it possible to address the identification of virtually any sample in the liquid, solid or gaseous state, solving challenging problems impossible to handle when using old dispersive spectrometers and traditional sampling techniques. Technological advances have also made it possible to transfer infrared spectroscopy instrumentation from the laboratory to on-site and *on-line* applications, by moving from benchtop instruments to portable miniature-type spectrometers[76,125,126]. Applications of these portable IR equipment are very diverse: study of air emissions[127], analysis of plastics[128], forensic science[129,130] and analysis of food and beverages[131,132]. IR transmission methods are the most traditional, but more modern methods based on light reflectance are gaining special attention. These are subdivided into attenuated total reflectance, diffuse reflectance and specular reflectance methods[133].

1.3.5.1. Transmission methods

Transmission is the oldest infrared method. It is based on the absorption of infrared radiation at specific frequencies as it passes through a sample. To analyse liquid solutions using this approach, the liquid is poured into a transmission cell of fixed or variable pathlengths. For the later, polytetrafluoroethylene (PTFE) spacers allow the cell to be used for various pathlengths, incorporating a mechanism for a continuous adjustment of the pathlength. Because of the high absorption of water in the transmission mode, alternative solvents should be considered. For solids, depending on the nature of the sample, three methods exist[116]:

- Alkali halide discs: the sample is grounded and mixed with an alkali halide powder. The most commonly used alkali halide is potassium bromide (KBr), because it does not produce a signal in the mid-infrared region. The mixture is sintered with pressure to produce a clear transparent disc.

- Mulls: the sample is ground and suspended in one or two drops of a mulling agent. Then, further grinding is applied until a smooth paste is obtained. Nujol (liquid paraffin) is the most commonly used mulling agent.
- Films: the sample is dissolved using a solvent capable of producing a uniform film. The resulting solution is poured onto a glass or metal plate and spread to a uniform thickness. The solvent is then evaporated and the film is stripped from the plate.

Gaseous samples are analysed using a gas cell, which is usually filled by flushing or from a gas line. The walls of the cell are of glass or brass. The size of the sample compartment is limited and a multi-reflection gas cell is necessary to produce longer pathlengths.

The use of vibrational spectroscopy in the transmission mode is not suitable for the *on-line* measurement of food and beverages as it involves a laborious sample preparation before analysis. Reflectance methods are preferred instead[116].

1.3.5.2. Reflectance methods

Reflectance methods arose as an alternative for samples that are difficult to analyse by transmission methods. There are three types of reflectance modes, depending on how the beam interacts with the sample[133].

1.3.5.2.1. Attenuated Total Reflectance

Attenuated Total Reflectance (ATR) is based on the principle of total internal reflection of an evanescent wave, that is, when the angle of incidence at the interface between the sample and a crystal is greater than the critical angle, the infrared beam entering the crystal will undergo total internal reflection. The critical angle is a function of the refractive indices of the two surfaces. The beam will penetrate a fraction of a wavelength beyond the reflecting surface and, when a material that absorbs radiation is placed in close contact with the reflecting surface, the beam will lose energy at the wavelength where the material absorbs. The resultant attenuated radiation will give rise to the absorption spectral characteristics of the sample. The amount of radiation that will penetrate the sample, known as depth of penetration (d_p), is given by:

$$d_P = (\lambda/n_1)/\{2\pi[\sin\theta - (n_1 / n_2)^2]^{1/2}\} \qquad \textbf{Equation 3}$$

where λ — wavelength, n_1 — refractive index of the sample, n_2 — refractive index of the crystal, θ —angle of incident radiation. As implied by the previous equation, the absorbing

path length of the infrared beam (0.5–5 μm) can be modified by angle of incident radiation, and it can be enhanced by applying multiple reflections in the sample (Figure 10)[78].

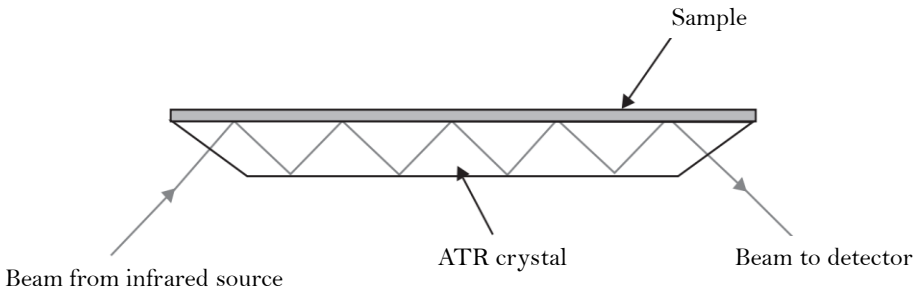


Figure 10. Schematic diagram of the attenuated total reflectance sampling technique. Adapted from Fagan and O'Donnell, 2008[134].

The crystals used in ATR, which act as internal reflection elements, are made from materials with low solubility in water and very high refractive index. The refractive index of the crystal must be significantly larger than that of the sample. The most common materials are zinc selenide (ZnSe), germanium (Ge) and thallium–iodide (KRS-5). However, the most suitable material is diamond, the high cost of which is compensated by its very long shelf-life and stability. Both liquid and solid samples can be analysed using ATR. In addition, a flow-through ATR cell can be installed, allowing the continuous flow of solutions through the cell[116].

The ATR method is gaining popularity because it solves important aspects in IR analysis: it drastically reduces sample preparation, the required sample volume is minimum and it increases spectral reproducibility[117].

Taking into account the above described advantages together with the analysis simplicity and rapidness, this technique was the one used in this thesis for wine alcoholic fermentation monitoring.

1.3.5.2.2. Specular reflectance

Specular reflection (also known as surface reflection) is based on external reflection principles, in which the incident radiation is focused onto the sample and the reflected radiation from the sample is measured. Specular reflectance occurs when the angle of reflection is equal to the angle of incidence. The sample must be reflective or attached to a reflective surface. This method is very useful for the study of solids with flat surfaces when

the sample cannot be melted, dissolved or finely ground. The sample is simply placed on a flat surface and measured. Liquid samples can also be measured using a PTFE cell[135].

The amount of light reflected depends on the angle of incidence, the refractive index, the surface roughness and the absorption properties of the sample. Grazing angle sampling accessories allow sample measurements at different angles of incidence. The resulting spectrum must be corrected using mathematical methods to get the traditional transmission spectrum[116].

1.3.5.2.3. Diffuse Reflectance

This method is commonly referred as Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) and is also based on the external reflectance of infrared radiation. When the incident infrared light is directed to the sample, it penetrates one or more particles and it is reflected in all directions giving rise to diffusely scattered light[117]. This method is especially suitable to analyse powdered samples mixed with KBr and rough surface solids. The DRIFTS cell reflects radiation to the sample and collects the energy reflected back over a large angle. The obtained reflectance spectrum relates the sample concentration to the scattered radiation intensity according to the Kubelka-Munk theory. Figure 11 shows the difference between specular reflectance and diffuse reflectance from a solid surface.

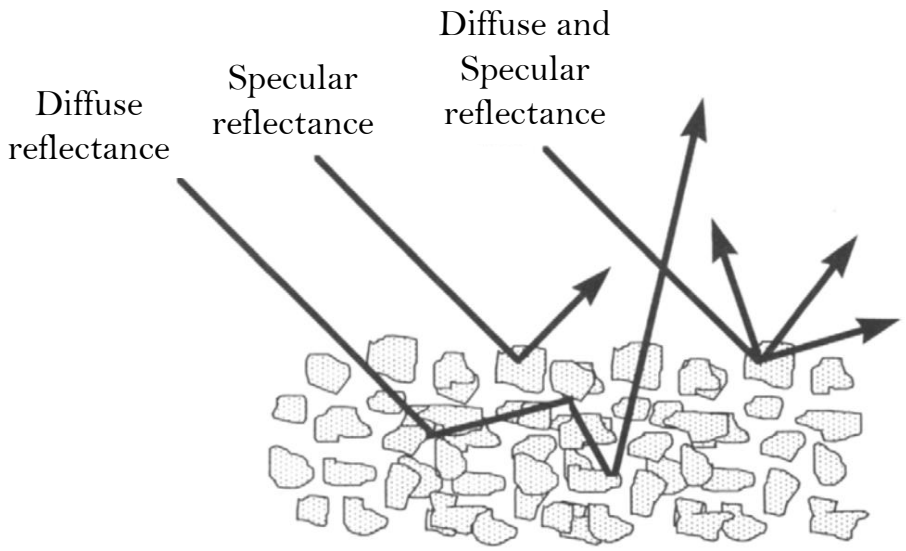


Figure 11. Representation of diffuse and specular reflectance in a powder sample. A single beam (right) can have a specular and a diffuse component at the same time. Adapted from Coates, 1998[136].

1.4. Multivariate data analysis: chemometrics

The first goal in the implementation of any PAT methodology is to get process data, which may include operating variables and process parameters. To acquire these data, the available tools have evolved from those that predominantly take simple univariate measurements, such as pH, temperature, or density, to those that measure biological, chemical and/or physical attributes using fast analytical instruments such as IR spectrometers.

Such analytical instruments provide a large amount of data (in the form of spectral signals) in a short period of time. In order to properly process these data and extract the desired information, multivariate data analysis (MVDA) techniques are necessary. MVDA techniques are typically used in areas such as consumer and market research, quality control and quality assurance, process optimization and control, research and development. In chemical sciences, these techniques are known as "chemometrics", and include all the appropriate mathematical algorithms that support analytical instruments providing large amounts of data in chemical measurements and chemical processes[137].

1.4.1. Multivariate IR data

As mentioned in the previous section, IR spectra are complex combinations of vibrational absorptions from all the bonds contained in the molecules of a given sample. For this reason, these spectra are difficult to interpret and the use of MVDA techniques is needed to extract the relevant information. Typically, for the analysis of a single sample with IR, the instrument will provide a numeric value (absorbance, transmittance or reflectance) for each of the measured frequencies in the IR region (wavenumbers or wavelengths). As a result, when multiple spectra from different samples are measured, high amounts of data are obtained that can be gathered in a data matrix with dimensions IxJ, where I refers to the number of samples (spectra) and J refers to the number of frequencies (wavenumbers). Furthermore, when considering a process, spectra are obtained at different time points, and time can be considered as a third dimension (K). The resulting matrix (Figure 12) has then three dimensions (IxJxK), which is the typical structure of a MVDA data matrix for PAT analyses[138].

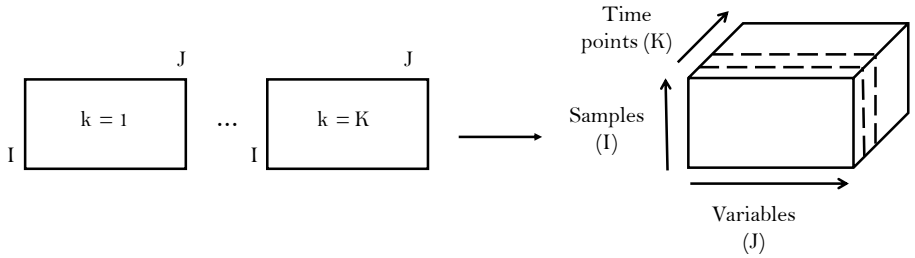


Figure 12. Representation of the three-dimensional multivariate data arrangement (right) of several data matrices containing the I samples of J variables for each K time point. Adapted from Figuereido, 2016[139].

1.4.2. Data preprocessing

Data preprocessing is a common step before multivariate data modelling. It involves the mathematical transformations of the raw data (i.e. in case of spectral data, the spectra resulting directly from the analytical instrument), in order to obtain meaningful qualitative or quantitative analytical models. Data preprocessing, can substantially enhance the quality of analytical data, as it minimizes the effects of instrumental or random noise, spectral artefacts, reduces sources of spectral systematic variability and improves selectivity. When the signal-to-noise (S/N) ratio of a given spectrum is very small, the application of preprocessing methods is especially important to maximize spectral differences.

For spectroscopic data, preprocessing strategies must be carefully studied to avoid losing useful information. Some knowledge is required about the nature of the variability present in the spectra and how the different preprocessing algorithms work[140].

Systematic variability in IR spectra is mainly due to scattering effects, baseline shifts, trending and noise (Figure 13). Depending on the source of variability arising from spectroscopic data, different preprocessing techniques can be applied to correct or minimize such problems[141]:

- **Noise:** is common to almost any analytical technique. In spectroscopy, it is related to small variations in absorbance that are not expected to be significant. The most common algorithm to reduce noise is the Savitzky-Golay smoothing, which is based on fitting polynomials to several data windows (one polynomial per window) with the same number of data points in each window[142].

- **Baseline effects**: the most common baseline effects in spectroscopy are offset and slope (Figure 13). In IR spectra, these effects lead to signals in which the baseline has shifted vertically (offset) or shows any type of baseline slope. Baseline effects

can be corrected by estimating a common baseline for all spectra, which is subtracted from each of the measured spectra. The most common baseline correction methods for spectroscopic data are Detrending and Multiplicative Scatter Correction (MSC). In detrending, one fits a polynomial of a fixed degree to the spectrum and subsequently subtracts this polynomial from the spectrum. In MSC, correction coefficients are estimated and a reference spectrum, which is usually the average spectrum of the calibration set, are used to build a corrected spectrum. First-order derivatives can also remove constant additive offsets, and second-order derivatives can remove baseline slopes. Second-order derivative, despite removing both offset and slope, has a major disadvantage, and is that it can increase the random noise of the signal.

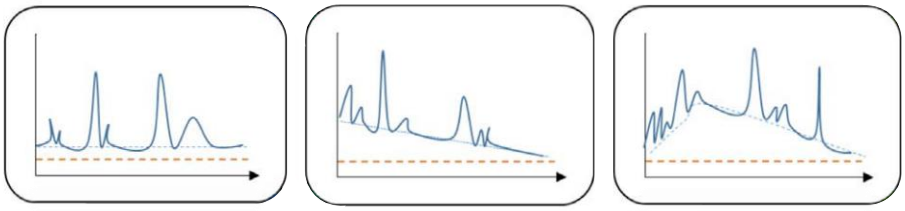


Figure 13. Most common baseline effects found in IR spectra. Baseline offset (left) represents a vertical shift at a constant value. Baseline slope (centre) is characterised by shifted baseline at an inconsistent value with an increase or decrease. Baseline curvature (right) is a special case of baseline slope in which baseline shifted at an inconsistent value which resembles a curve shape. Adapted from Chuen Lee *et al.*, 2017[143].

- **Scattering effects:** IR spectra are affected by light scattering. Light scattering occurs because the size of the particles in the sample has at least one dimension that is roughly the same magnitude than the spectroscopic wavelengths. These effects usually appear in IR spectra as a drift in the signal. This effect is greater for high wavenumbers, and can be due to scattering of the infrared light at the surface or inner part of the sample. The most common preprocessing methods to correct scattering effects are Standard Normal Variate (SNV) and Multiplicative Signal Correction (MSC). The first method subtracts the spectrum mean from each spectral variable and subsequently divides that value by the standard deviation of the spectrum (i.e. the estimated scatter constant). As described previously, MSC estimates the coefficient describing the scattering by regressing the spectrum to correct onto a reference. It must be noted, though, that SNV has to be taken with caution, as it "shifts" informative regions along the signal range, making it difficult to interpret the spectra[144]. There exist newly developed algorithms to avoid this problem, such as Variable Sorting for Normalization (VSN), in which the variables

constituting the signals are suitably weighted before calculation of the mean and the standard deviation used in the SNV transform[145,146].

Derivatives are not only useful for minimizing baseline effects, but also to reduce noise (when used in combination with smoothing) and enhance slight spectral differences between nearly identical samples. More recent preprocessing techniques include Orthogonal Signal Correction (OSC) or External Parameter Orthogonalisation (EPO). OSC aims at removing all the variations in the spectra that are not related to a response variable. EPO specifically removes information related to the variation of an external parameter that negatively influences the response variable[147–149].

In many cases, mean centering and standardization are applied to the data after preprocessing. In mean centering, the mean of all the absorbance values from each variable is calculated and subtracted from each single value. The mean for all values is set to zero, which means that negative values will appear for some of the values. Mean centering is useful to emphasize the differences between samples. Standardization will make each column have the same "size", as each single value is divided by the standard deviation of the column. The joint application of mean centering and standardization is referred as autoscaling [150,151].

The choice of the most suitable combination of preprocessing techniques depends on the type of spectral signal and the nature of the application. Generally, there are three types of preprocessing approaches, which are commonly found in the literature:

- Trial and error, in which the decision on the best strategy is based on the best performing one according to the goal of data analysis.
- Visual inspection where, for each pre-processing strategy of interest, the pre-processed data are inspected by simply looking at these data and checking if any artefacts are still visible.
- Assessment of pre-processed data by data quality parameters, which aims at quantifying the presence of artefacts in the data[141].

1.4.3. Variable selection

In chemometrics, it is important that variations in the data that are not related to the signals of interest are removed before building a chemometric model. Typically, chemometric algorithms are able to disregard this unimportant sources of variation. However, a variable selection step may be required in some cases, as part of the preprocessing strategy before

applying any MVDA method. This step aims at selecting specific wavelengths out of the whole spectra, that are relevant to the signal of interest and that better define the sample class or better correlate with the contents of analytes or properties of interest, removing non informative regions that could be hampering the model performance. The main idea is to remove those variables not contributing to improve the performance of the overall model. The procedure is normally conducted after some preprocessing of the raw spectra set. These methods are commonly applied when building prediction models for quantification purposes, such as Partial Least Squares Regression (PLSR), which will be discussed in the following section[152,153].

There exist three types of optimization methods for variable selection:

- **Filter methods**: these methods are applied after building a chemometric model, and are generally implemented defining a ranking criterion and applying a threshold arranging variables into importance measures. The most well-known filtering methods include the Variable Importance in Projection (VIP) and the Selectivity Ratio (SR). Based on the outcome of such filter methods, variables can be selected by, e.g. setting a threshold on the value of the variable importance measure[154].

- **Wrapper methods**: in this case, variable selection is accomplished by iteratively removing variables from the data and refitting a chemometric model on the reduced data. Model performance is evaluated to choose the best subset of variables. Some wrapper methods are: uninformative variable elimination in PLS (UVE-PLS), interval PLS (iPLS), genetic algorithm (GA) and Predictive Property-Ranked Variable (PPRV)[155].

- **Embedded methods:** in this type of methods, the variable selection is integrated in model building, so that the computation time required for reclassifying different subsets (as done in the wrapper methods) is reduced. The best subset of variables is assessed as part of the training process to build the model. Some of the typical methods of this category are based on the introduction of sparsity, e.g. in weight vectors, regression coefficients (PLS), loadings (PCA) or canonical vectors (LDA)[156].

Usually, model performance and interpretation is expected to improve through a correct combination of preprocessing and variable selection techniques, by highlighting the chemically relevant variables. It should be noted that, above all, variable selection should

be performed with a priori knowledge of the researcher about the process and the IR spectrum, and that it is possible to build a multivariate model by restricting the analysis to a particular IR spectral region containing specific chemical information[143].

1.4.4. Exploratory data analysis

Once the data set has been preprocessed, the first step in MVDA is exploring the data variability to get some preliminary information. This type of analysis is often called unsupervised, because the analysis is performed without taking into account any label or category. The most common method for data exploration is Principal Component Analysis (PCA). This method allows the identification of patterns in the data, including similarities and differences (for samples or variables), outliers and trends between data variables or samples[157]. PCA is based on dimensionality reduction, by transforming the original correlated variables into a fewer number of independent variables, while preserving the maximum variance information as possible. These new uncorrelated variables are called Principal Components (PCs), and they are orthogonal to each other. PCs are a linear combination of the original variables, and they are built by detecting the directions in space (the original space has *J* dimensions, being *J* the number of initial variables) involving the largest variability in the data. The first PC detects the direction with the largest variability, the second PC detects the direction with the second largest variability and is orthogonal to the first PC, and so on. Therefore, each PC is orthogonal to its previous PC, and detects the maximum variability which has not been explained by the previous PCs[158]. PCs represent the coordinates of the new space created by the PCA model. In practice, the first few components are the ones retained by a PCA model, as those include the useful information. Each PC is defined by the product of two vectors: the score vector ($\mathbf{t}_a$), and the loading vector ($\mathbf{p}_a$), where *a* is the number of a given PC. Data decomposition in fewer variables can be described by the following equation[159]:

$$\mathbf{X} = \mathbf{t}_1 \cdot \mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2 \cdot \mathbf{p}_2^{\mathrm{T}} + \cdots + \mathbf{t}_A \cdot \mathbf{p}_A^{\mathrm{T}} + \mathbf{E} \qquad \textbf{Equation 4}$$

where $\mathbf{X}$ (IxJ) is the original data matrix, $\mathbf{t}_a$ is the score vector for the *a*th component, $\mathbf{p}_a$ is the loading vector for the *a*th component and $\mathbf{E}$ (IxJ) is the residual matrix containing the variation not modelled. A schematic representation of the PCA matrix decomposition is shown in Figure 14. The goal of PCA is to explain as much as possible the total sum of squares of $\mathbf{X}$ with the minimal number of PCs[159]. The calculated scores and loadings can be plotted in simple 1D or 2D graphs, allowing the interpretation of the whole data space in

an easy and visual way. Scores represent the projection of the samples into the new system of coordinates (the $\mathbf{p}_a$ loadings) and they explain relationships between samples. In turn, loadings represent the weight coefficients of the original variables with respect to the PCs and they explain relationship between variables and PCs, and variables themselves.

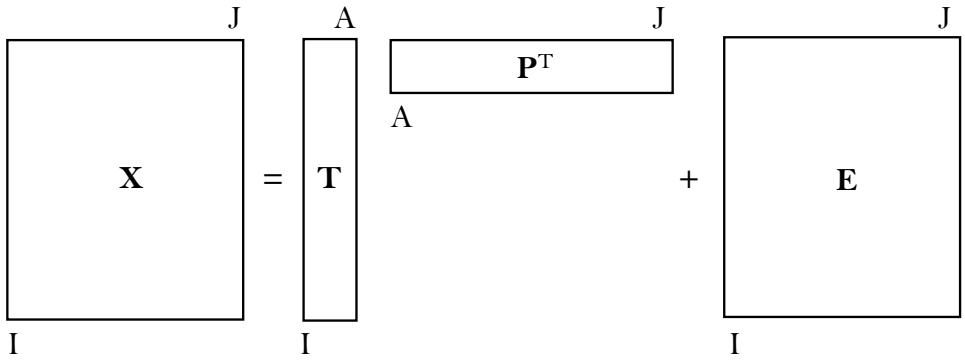


Figure 14. PCA decomposition in matrix form. The original matrix (**X**) is decomposed in the product of scores (**T**) and loadings ($\mathbf{P}^T$) plus the residual matrix (**E**).

In PCA, it is common to calculate two statistical parameters in order to evaluate the performance of the model: Hotelling T^2 and Q statistics. These two statistics are of great importance when determining if a sample is an outlier, that is, a sample showing unusual features. Detection and removal of outliers is necessary before building any MVDA model, as failing to do so can dramatically affect the performance of the model. The Hotelling T^2 statistic can be seen as an extension of the univariate *t*-test and can be applied to the scores of a PCA. It defines the statistical boundaries assuming a normal distribution of scores in the PCA, and it represents the distance of a sample to the center of the PCA model. Hotelling T^2 for a given *i* sample is calculated as:

$$T_i^2 = \frac{\mathbf{t}_i^T(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{t}_i}{\mathbf{I-1}} \qquad \textbf{Equation 5}$$

where **T** (IxA) is the score matrix of the calibration samples and $\mathbf{t}_i$ (Ax1) is a vector containing the A scores of the projected *i*th sample in the PCA space. A confidence limit can be calculated for the Hotelling T^2, assuming that scores are normally this distributed (Equation 6). This limit can be used to control if some of the samples are showing unusual score values[160].

$$T^2_{i\,(\mathbf{I,A})} = \frac{\mathbf{A(I-1)}}{\mathbf{I-A}} F_{\mathbf{A,I-A},\alpha} \qquad \textbf{Equation 6}$$

where A is the number of principal components, I the number of samples, and α the significance level of the test. The Q statistic is simply the sum of the squared residuals of each sample. Residuals can be defined as the squared difference between the measured values and the values calculated from the PCA (Equation 7) and it represents the perpendicular distance of a sample to the space of the PCA defined by the selected A components.

$$Q_i = \mathbf{e}_i\mathbf{e}_i^T = \mathbf{x}_i(\mathbf{I} - \mathbf{P}_A\mathbf{P}_A^T)\mathbf{x}_i^T \qquad \textbf{Equation 7}$$

where $\mathbf{e}_i$ is the ith row of $\mathbf{E}$, $\mathbf{P}_A$ is the matrix of the A loadings vectors retained in the PCA model (where each vector is a column of $\mathbf{P}_A$) and $\mathbf{I}$ is the identity matrix of appropriate size (J x J)[161]. As for the Hotelling T^2 statistic, a confidence limit can be calculated for the Q-residual statistic[162]. The Q statistic represents the part of the information of the data that cannot be explained by the PCs in the PCA space[151]. Both statistics can be plotted together, in what is called "influence plot", which is a very useful tool to look for samples that are not well described by the PCA model.

1.4.4.1. ASCA: a new exploratory tool

When studying a process, the influence of several factors is usually considered in the Design of Experiments (DoE). However, when dealing with multivariate data, it is very difficult to discern which part of the variables is directly influenced by the experiment factor and which is not. ANOVA-Simultaneous Component Analysis (ASCA) is a quite recent exploratory method in chemometrics, which aims at evaluating the significance of one or more experimental factors. Hence, it can be considered a direct generalization of analysis of variance (ANOVA) for univariate data, but applied in a multivariate way[163].

In the case of two known experimental factors, ASCA decomposes the centered data matrix ($\mathbf{X}_c$) according to:

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = \mathbf{X}_{\text{factor 1}} + \mathbf{X}_{\text{factor 2}} + \mathbf{X}_{\text{factor 1 x factor 2}} + \mathbf{E} \qquad \textbf{Equation 8}$$

where $\mathbf{X}$ is the original data matrix, $\mathbf{1}$ is a vector of ones, $\mathbf{m}^T$ is the mean of all the observations, $\mathbf{X}_{\text{factor 1}}$ and $\mathbf{X}_{\text{factor 2}}$ are the matrices representing the effects of each one of the experimental factors, $\mathbf{X}_{\text{factor 1 x factor 2}}$ is the interaction matrix of the factors and $\mathbf{E}$ is the residual matrix. Each matrix is centered and contains the mean profiles of the samples corresponding to each factor or interaction level. As an example, if factor 1 has two levels

of 20 observations each, 20 observations will contain the average profile for the first level of factor 1, and the same will happen with level 2[164]. The interaction matrix is calculated after the subtraction of the main effect matrices. Since all the effect matrices are centered, the magnitude of the effects can be evaluated as the sum of the matrix elements. Given a factor i:

$$\mathbf{SSQ}_{\text{factor } i} = ||\mathbf{X}_{\text{factor } i}||^2 \qquad \textbf{Equation 9}$$

To evaluate if the effect of a particular factor or interaction is statistically significant, the value of the sum of squares of the corresponding matrix is compared to its distribution under the null hypothesis, as evaluated non-parametrically by a permutation test[165]. The permutation test allows the calculation of a *p*-value for samples that do not necessarily meet the conditions of normality. Given a factor i and its associated matrix $\mathbf{X}_i$, the *p*-value is calculated as:

$$p - \text{value}(\mathbf{X}_i) = \frac{\text{nbr}(\mathbf{SSQ}(\mathbf{X}_{i,\text{perm}} \geq \mathbf{SSQ}(\mathbf{X}_i)))}{\text{k}} \qquad \textbf{Equation 10}$$

where "nbr" is the number of occurrences, k is the number of permutations, $\mathbf{X}_{i,\text{perm}}$ is the matrix obtained after a random row permutation. Thus, the *p*-value relates the number of cases where the variance of the studied factor is lower than the variances resulting from the permutations. In this way, the effect of the studied factor is compared to its distribution under the null hypothesis as estimated by the permutations[166,167].

Then, a bilinear decomposition of each matrix is performed using Simultaneous Component Analysis (SCA). In the context of ASCA (under the constraints of ANOVA), this reduces to PCA, as the goal is to model the variability linked to each of the factors. Hence, each matrix from Equation 8 can be decomposed as:

$$\mathbf{X}_{\text{factor } i} = \mathbf{T}_{\text{factor } i} \cdot \mathbf{P}^{\text{T}}_{\text{factor } i} + \mathbf{E}_{\text{factor } i} \qquad \textbf{Equation 11}$$

where $\mathbf{T}_{\text{factor } i}$ is the score matrix, $\mathbf{P}^{\text{T}}{}_{\text{factor } i}$ is the loading matrix and $\mathbf{E}_{\text{factor } i}$ is the residual matrix of the ith partitioned matrix in Equation 8. Reducing dimensionality enables a better interpretation and visualization of the data considering each experimental factor or interaction separately. The loadings for factor i define a subspace spanned by $\mathbf{X}_{\text{factor } i}$, highlighting the spectral directions related to the factor under study. In turn, the scores for factor i are the new coordinates of the observations on the PCs of the model[166,168,169].

1.4.5. Predictive analysis

Predictive methods focus on solving, using multivariate data, the regression problem, one of the most common data-analytical problems in science and technology. These methods aim at determining how to model one or several dependent variables (**Y** responses), by means of a set of predictor variables contained in a data matrix **X**. The **Y** responses are usually concentrations of chemical compounds or key parameter values (i.e. density)[170].

The gold-standard for multivariate regression in chemometrics is Partial Least Squares Regression (PLSR). PLSR relates two data matrices (**X** and **Y**) through a linear multivariate model. In PLSR, instead of finding those directions which maximize data variability in the X-space, the algorithm tries to find those directions in the X and Y-spaces corresponding to the maximum covariance between **X** and **Y** to get their best correlation, while explaining as much as possible the variability in **X**[171].

PLSR seeks to build a linear regression equation between the scores of the predictor variables and the scores of the dependent variables. First, both **X** and **Y** matrices are decomposed creating principal directions that describe the maximum variability of **X** while considering the maximum covariance between the scores of **X** and **Y**. The **X** scores estimate the linear combination of the x variables, with weight coefficient **W***:

$$\mathbf{T} = \mathbf{X} \cdot \mathbf{W}^* \qquad \textbf{Equation 12}$$

The weight **W** can be transformed to **W***, which is directly related to **X**. From Equation 12, **W*** is defined as:

$$\mathbf{W}^* = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \qquad \textbf{Equation 13}$$

The PLSR model consists of an outer and an inner relation. The outer relations describe the **X** and **Y** matrices individually, and are given by:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad \textbf{Equation 14}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^T + \mathbf{F} \qquad \textbf{Equation 15}$$

where $\mathbf{P}^T$ and $\mathbf{C}^T$ are loading matrices of the **X** and **Y** spaces, respectively. **T** is the score matrix of **X** and **U** the score matrix of **Y**. **E** and **F** are the residual matrices of **X** and **Y**,

respectively. **X** scores (**T**) are good predictors for the **Y** variables. The inner relation between X and Y-scores (the new variables) can be defined by:

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^{\mathrm{T}} + \mathbf{G} \qquad \textbf{Equation 16}$$

By combining Equations 12 and 16, we can redefine **Y** as:

$$\mathbf{Y} = \mathbf{X}\mathbf{W}^{*}\mathbf{C}^{\mathrm{T}} + \mathbf{G} = \mathbf{X}\mathbf{B} + \mathbf{G} \qquad \textbf{Equation 17}$$

where

$$\mathbf{B} = \mathbf{W}^{*}\mathbf{C}^{\mathrm{T}} \qquad \textbf{Equation 18}$$

B are the coefficients of the PLSR model and **G** is the residual matrix. The prediction of y-variables for new samples is calculated using Equation 17 [172].

Scores and loadings have the same properties described for PCA. However, in PLSR the loadings are linear combinations of the original variables, which are known as latent variables (LVs) and explain the variability of **X** that most correlates with **Y**. Loadings can be interpreted as the influence of each variable on each LV, while scores represent the projection of the samples in the LV space[66].

1.4.5.1. Calibration and validation steps

Predictive analysis consists of two steps (Figure 15). First, a PLS model which best correlates **X** and **Y** is built and optimised, based on the data available. This is called calibration (or training) step. Once the model has been created, the regression coefficients obtained are used for the second step, in which new external data ($\mathbf{X}_{\text{test}}$), which were not used in the training step are presented to the model, and the values for **Y** are predicted. The data set used in this step is called test (or validation) set. When the number of samples is limited, it is common to apply a cross-validation (CV) strategy. Cross-validation involves selecting blocks of samples from the original training set. These samples are left-out, the model is built with the remaining samples and the left-out samples are used for subsequent prediction, to test the performance of the model. The process is repeated for consecutive blocks of samples. When only one sample is left-out at a time the method is called leave-one-out cross validation.

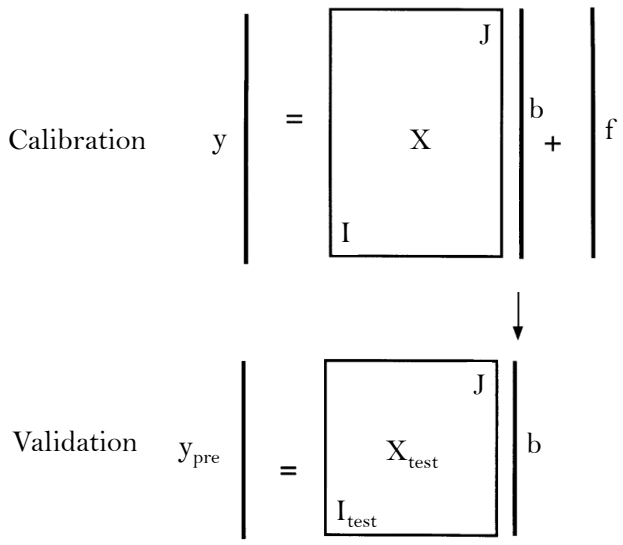


Figure 15. Schematic representation of the calibration and validation processes using PLSR. Adapted from Geladi, 2003[159].

When selecting the number of optimal LVs in a PLSR model, special precaution must be taken to avoid overfitting, which means that the model may show a great fitting ability for a specific set of data, but when new data are used to predict the Y variables using the same PLSR model, the model fails at providing accurate results.

To determine the optimal number of LVs, the root mean error of prediction (RMSEP) is typically used (Equation 19), which serves to evaluate the performance of a PLSR model.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad \textbf{Equation 19}$$

y_i and $\hat{y}_i$ are the measured and predicted values for the ith sample, and n is the total number of samples in the validation set. When cross-validation is performed, the root mean squared error of cross-validation (RMSECV) is calculated (Equation 20):

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{i,\text{CV}})^2}{n}} \qquad \textbf{Equation 20}$$

where y_i and $\hat{y}_{i,CV}$ are the measured and predicted values for the ith sample, and n is the total number of cross-validation samples[170].

1.4.6. Classification analysis

Multivariate classification methods in chemometrics (also known as supervised pattern recognition methods) aim at classifying samples using the information provided by a set of measurements. Classification models are based on membership recognition of each sample to its appropriate class. Once a classification model has been developed, unknown samples can be classified. A mathematical relationship between a qualitative variable (i.e. categories or classes) and descriptive variables (e.g. chemical measurements, spectra) is established[173].

There are two different approaches for classification analysis: discriminant and class-modelling methods. When using a discriminant method, a sample is always classified into a category. In class-modelling, a sample could be assigned to or rejected from one or more classes. One of the most common classification methods is Partial Least Squares Regression Discriminant Analysis (PLS-DA), which was originally proposed to an extension of PLSR[155].

In PLS-DA, the class vector **y** containing the membership of the samples is transformed into a dummy matrix **Y** (Figure 16), of dimensions IxC, where I is the number of samples and C is the number of classes in the dataset. Each y_{ic} value in the dummy matrix represents the membership of the *i*th sample in the *c*th class expressed in binary code (1 if it belongs to a given class or 0 if not)[174].

$$\text{Class Vector} = \begin{vmatrix} s_1 \\ s_2 \\ \cdots \\ s_i \end{vmatrix} \quad \text{Dummy matrix (Y)} = \begin{vmatrix} 1_{11} & 0_{12} & \cdots & 0_{1c} \\ 0_{21} & 1_{22} & \cdots & 0_{2c} \\ & & \cdots & \\ 0_{i1} & 0_{i2} & \cdots & 1_{ic} \end{vmatrix}$$

Figure 16. Class vector containing the membership number for each class (left) and subsequent transformation into a dummy matrix of binary values (right). Adapted from Pomerantsev and Rodionov, 2018[175].

Then, the PLSR model is built in the usual way, and estimated values of **Y** around 0 and 1 are obtained for each sample and class. Although the estimated class values will not be exactly zero or one, if a predicted *y* value for a given class is closer to zero than to one, then it will not likely belong to that class, while a value closer to one would indicate the opposite.

To make a class assignment, a probability is calculated for each class and samples are therefore classified by choosing the class that has the highest probability. In this way, samples are always classified to one class. A decision threshold may also be determined based on the Bayes theorem, assuming that the estimated values follow a normal distribution, which should be comparable to what will be observed for future samples. The threshold is usually selected at the point where the number of false positives and false negatives is minimized, but other methods can be applied on order to avoid the limitation of using an arbitrary threshold for assignment between the classes[173,176].

As in PLSR, calibration and test steps are also needed to validate a PLS-DA model. The calibration set should have a balanced proportion of representative samples for each class to build the calibration model, and external test samples rather than CV samples should be preferred to evaluate the predictive ability of the calibrated model[173].

1.4.7. Resolution analysis

Multivariate Curve Resolution (MCR) comprises a series of mathematical methods that aim at resolving the composition of a mixture into its pure components. One of the most important methods for resolution analysis is Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS). MCR decomposes the data matrix into a bilinear model, in a similar way as PCA, but it also adds a series of constrains to the profiles of the components that give the mathematical solution a chemical meaning. MCR methods can be defined as the multilinear extension of Lambert-Beer's law[177]. Beer's law is the fundamental law of quantitative spectroscopy that relates the intensities of the incident and transmitted IR radiation to the concentration of the analyte. Lambert-Beer's law can be described as:

$$A = \log_{10}\left(\frac{I_0}{I}\right) = \varepsilon cl \qquad \textbf{Equation 21}$$

where A is the absorbance of the sample, I_0 is the intensity of the light entering the sample, I is the intensity of the light transmitted by the sample, ε is the molar absorptivity, c the concentration and l the path length of the sample[111].

In MCR, the original data matrix (**D**), with dimensions' IxJ (being I the number of observations (i.e. spectra over time) and J the number of variables (i.e. wavenumbers) is decomposed as follows:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \qquad \textbf{Equation 22}$$

where **C** (I x F) is the matrix of concentration profiles, which represents the variation in weight (abundance) of those particular compounds influencing the spectral signals. *d* is the number of relevant components found by the algorithm to resolve the mixture and *m* are the observations over time along the row direction in the data set. $\mathbf{S}^T$ (F x J) is the matrix of pure spectral profiles of dimensions, being *n* the number of variables). Finally, **E** (I x J) is the residual matrix. An schematic overview of MCR the decomposition is shown in Figure 17[178].

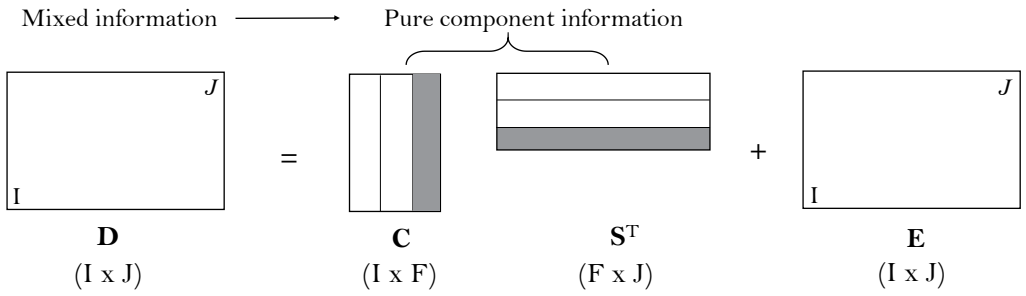


Figure 17. Schematic representation of the MCR decomposition. Each column ($\mathbf{c}_j$) in matrix **C** represents the concentration profile of a component, while each row in $\mathbf{S}^T$ ($\mathbf{s}_j$) represents the spectral profile of the pure component. Adapted from De Juan and Tauler, 2003[178].

Before applying MCR-ALS, the number of optimal components is selected by applying the PCA algorithm. The reduction of dimensionality in the original matrix is done using the Single Value Decomposition (SVD) method. Then, it is possible to include spectral information in the form of initial estimates of concentration or response profiles if we have information about the components present in the mixture[177].

The optimization of MCR matrices with ALS involves an iterative approach, which starts from initial estimates of **C** or $\mathbf{S}^T$ that evolve until profiles with chemically meaningful shapes are obtained. The goal of the MCR-ALS iterations is to minimize as much as possible the residuals in the **E** matrix using least-squares and applying suitable constraints. Constraints involve the accommodation of external chemical information (chemical knowledge of the system) into the optimisation process, and they are necessary because the product $\mathbf{C}\cdot\mathbf{S}^T$ is subjected to rotational and intensity ambiguities[179].

The most common constraints for chemical systems are[180]:

- **Non-negativity of spectral profiles**: positive values for absorbance values are imposed, as negative absorbance values do not have chemical meaning.
- **Unimodality**: presence of only one maximum in each concentration profile.

- **Non-negativity of concentration profiles**: only positive values for concentrations are allowed, as real concentrations cannot be negative.
- **Closure**: mass balance conditions must be fulfilled for reaction-based systems.

A common criterion to stop the MCR-ALS algorithm is the percentage of Lack of Fit (%LOF). When the relative difference between two consecutive iterations is equal or less than 0.1%, the algorithm stops the calculation. The %LOF is defined by the following equation:

$$\text{lack of fit (\%)} = 100 \cdot \sqrt{\frac{\Sigma_{ij} e_{ij}^2}{\Sigma_{ij} d_{ij}^2}} \qquad \textbf{Equation 23}$$

were d_{ij} is an element of the experimental data matrix **D** and e_{ij} is the related residual value obtained from the difference between **D** and $\mathbf{CS}^T$ (the matrix product obtained by MCR-ALS)[181].

It is possible to augment the **D** matrix in a column-wise manner (Figure 18) to add more experiments into the original dataset. In this way, the $\mathbf{S}^T$ matrix will still be a single matrix containing the spectral profiles of the components in the different experiments and **C** will be composed by $\mathbf{C}_i$ submatrices, being i the number of individual experiments, Each $\mathbf{C}_i$ matrix will show the unique concentration profiles of each experiment that are independent to each other[177].

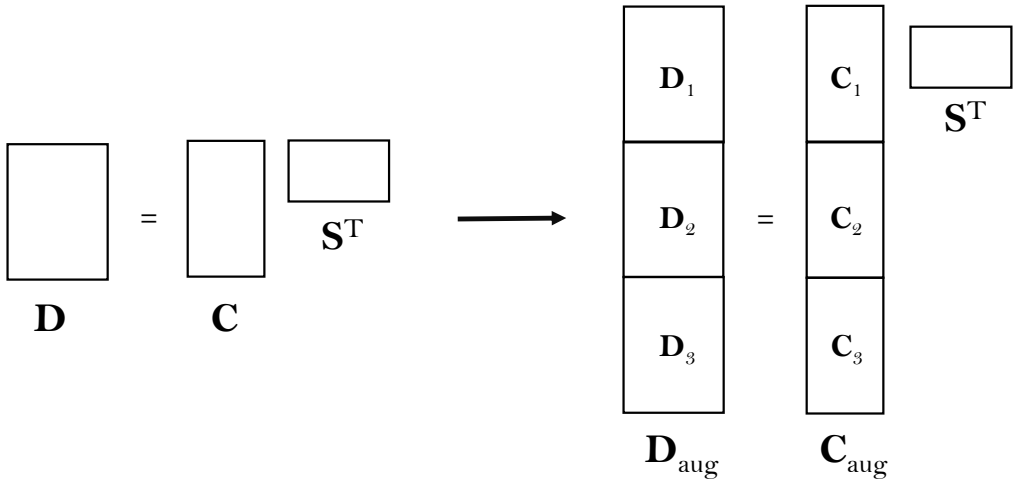


Figure 18. **D** matrix augmentation through the introduction of additional experiment data ($\mathbf{D}_i$). **C** matrix augments, being each experiment independent to each other, whereas $\mathbf{S}^T$ remains the same. Adapted from De Juan and Tauler, 2006[177].

Resolution methods are increasingly being applied to study a wide variety of complex processes in biochemistry and biophysics from spectroscopic measurements. NIR and ATR-MIR data from fermentation-related bioprocesses have been studied using MCR-ALS, allowing the determination of the concentration and spectral profiles of the main species involved (sugars, ethanol and biomass production) and other compounds such as acetate, ammonium and total phosphate[182–184].

1.4.8. Multivariate statistical process control

Statistical process control (SPC) refers to statistical methods used to monitor and improve the quality and productivity of manufacturing processes. The goal of any SPC scheme is to monitor the performance of a process over time and determine if the process behaves as expected or detect any unusual event that may have occurred during the process. Knowing the causes of an unusual event can significantly improve the process performance, as they can be corrected or revised[185].

Control charts are the graphical expression of a process performance over time, which help to identify the magnitude and type of variation present in an observation. The implementation of PCA-based Multivariate Statistical Control (MSPC) charts has improved the ability to detect certain out-of-control situations. When using MVDA for process monitoring and control, MSPC charts are based on dimensionality reduction methods, such as PCA. By reducing the number of variables, the number of charts is also reduced, allowing a better and easier interpretation. Another significant benefit of using MSPC charts is that, unlike univariate statistical process control, interactions between variables and the combined effect of changes can be observed on the charts, in addition to univariate information[186,187].

The classical way to build MSPC charts is defining the reduced space for normal operation conditions (NOC) samples, which is done by building a PCA model using reference samples (i.e. those samples meeting the desired quality expectations). This is done by registering as much data as possible from a process over time. For this NOC model, limits are defined for T^2 and Q statistics, and new samples are projected onto the model. Using T^2 and Q statistics rather than the PCA score plot allows assessing in a better way how a sample fits in the model.

At different stages of the process, if a projected sample falls inside the confidence limits of T^2 and Q, the sample is considered normal, whereas if the sample falls outside those T^2 and Q limits it is considered an outlier[188]. The contribution of each original variable to the T^2

and Q statistics can be inspected to determine the causes of an unusual behaviour, helping the operators to determine quickly and efficiently the source of a problem. Figure 19 shows a typical example of MSPC charts[189].

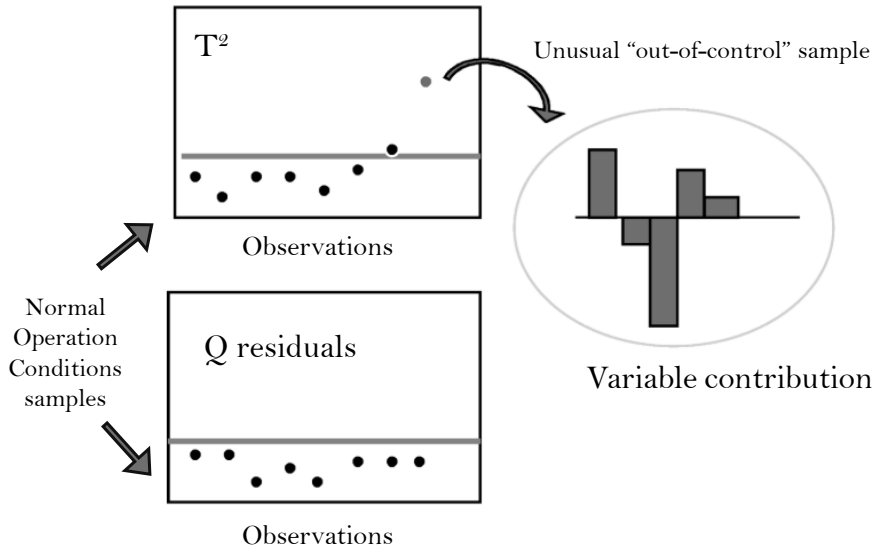


Figure 19. MSPC scheme at a glance. Samples below the threshold limit are considered under Normal Operation Conditions, whereas unusual samples will appear above the threshold. Adapted from Kourti, 2006[190].

Multivariate statistical methods have significantly increased the potential of using multivariate data for process quality control to meet the PAT initiative guidelines. The use of projection methods such as PCA has modernised the idea of SPC, allowing an effective monitoring of complex processes by only looking at a few multivariate control charts. These charts improve the fault detection capabilities as faults can be detected earlier than in univariate charts, were only few process variables can be controlled at a time. In contrast to univariate SPC charts, in which some problems can be missed if the variables remain within their expected univariate operation limits, problems in MSPC charts are detected as changes in the covariance structure of the process variables[190]. The application of MSPC charts based on PCA for process monitoring of fed-batch fermentations has been investigated in the past, suggesting the usefulness of this method for predicting the behaviour of future batches[191]. The use of MSPC has also been studied for monitoring petroleum refining or coffee roasting processes, allowing the detection of abnormal batches before the end of each process[192,193].

1.5. Wine fermentation process control

In the wine industry, even though traditional analytical methods are still the most reliable methods for the determination of quality control parameters, IR techniques are gaining ground due to all of their above-mentioned advantages. Replacing old methods with PAT tools will provide the wine industry with fast (even *real-time* in some cases) and simple measurements, allowing a better understanding of the process and giving the possibility to take corrective measures in time if some variables of the process are not under control. Chemometric methods have been widely used for the prediction of chemical parameters (mainly PLSR) and for classification (using discrimination or clustering techniques)[194].

1.5.1. IR spectra of wine in the mid-infrared region

In wine, the most important constituents are water and ethanol. These molecules show high absorption in the MIR region. For water, two broad peaks are found in the regions 3626-2970 cm^{-1} and 1716-1543 cm^{-1}. These peaks may hide characteristic bands of other minor constituents, such as organic acids and phenols. The most relevant region in wine MIR spectra is the fingerprinting region, which includes those vibrations corresponding to C-O, C-C, C-H and C-N bonds. This region provides important information regarding a large number of organic compounds such as sugars, alcohols and organic acids present in the sample, as their functional groups absorb in this region[195]. Figure 20 shows the main characteristic bands found in typical MIR spectra for different beverage samples.

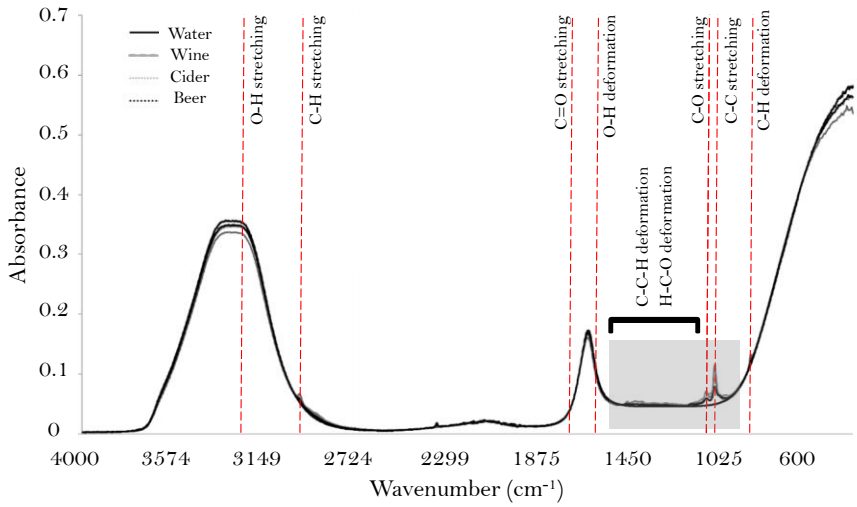


Figure 20. Average ATR-MIR spectra (4000–400 cm^{-1}) for beverage samples, pointing out some of the most important absorption peaks with their corresponding associated molecular vibrations. The fingerprint region is highlighted with a grey box. Adapted from Pearce *et al.*, 2006[196].

In the fingerprinting region, the bands are mostly influenced by the changes during alcoholic fermentation. In fact, the bands change rapidly as sugars transform into ethanol (Figure 21)[197].

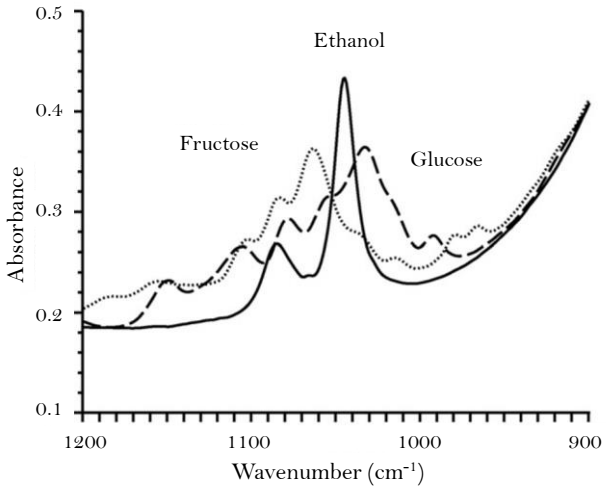


Figure 21. Detail of the spectral shapes for glucose, fructose and ethanol in the fingerprinting region of the MIR spectrum. From Wynne *et al.*, 2007[198].

Similarly, organic acids show important peaks at 1740 cm^{-1}, corresponding to the C=O bond stretching vibration of carboxylic acids. O-H bending of the carboxylic acid, C-O stretching of acid and alcohol (when present) and C-H bending vibrations in the fingerprinting region (around 1390 cm^{-1}) are also important bands for the determination of organic acids in must and wine samples. The close similarities among organic acid molecular structures causes spectroscopic interferences and difficult the quantification of individual organic acids in wine[199].

1.5.2. Analysis of must and wine with MIR and chemometrics

The analysis of must and wine with IR techniques involves the determination of a large number of wine parameters, including alcohol, volatile acidity, pH, tartaric acid, lactic acid, reducing sugars (mainly glucose and fructose), acetic acid, glycerol, anthocyanins and polyphenols[200]. It is by no surprise that the application of FT-IR spectrometers to measure the concentration of some of the most important molecules in must and wine has been extensively investigated in the last decades. Despite being considered very complex mixtures (especially fermentation samples, where the mixture is constantly evolving), the determination of a wide range of compounds in wine has been possible using FT-MIR and PLSR, with determination coefficients (R^2) higher than 0.9 for alcohol, reducing sugars

(glucose and fructose), total acidity and glycerol[201]. In a study with more than 250 samples of different types of wine (white dry, white sweet, blush, red and dry red premium), the standard error of calibration (SEC) was 0.106% v/v for ethanol, 0.308 g·100mL^{-1} for titratable acidity, 0.05 for pH and between 0.108 and 0.486 g·L^{-1} for tartaric, malic and lactic acids, depending on the type of wine[202].

The calibration and validation process for the determination of organic acids (tartaric acid, malic acid, succinic acid, and lactic acid) in red and white wines using FT-IR is, however, less accurate when the measurement range is below 0.6 g·L^{-1} [203]. In another study, the quantification accuracy for the major fermentation compounds in wine showed low Standard Errors of Prediction (SEP): volatile acidity (0.08 g·L^{-1}), ethanol (0.32%), glycerol (0.38 g·L^{-1}) and reducing sugars (0.56 g·L^{-1})[204]. For the determination of tannins, FT-IR combined with PLS only allows developing regression models to provide approximate quantitative values of individual anthocyanin concentrations in wine must samples. Tannin concentration in wine is in the order of milligrams, which could explain the high SEP values obtained through external validation (between 13 and 26%)[205]. Another study covering red wine samples from different vintages reported RMSEP values between 75 mg·L^{-1} and 115 mg·L^{-1} for tannins content, by testing different variable selection strategies[206]. Both studies concluded that the quantification of tannins is difficult due to interferences from spectral signals of other wine components. It is important to note that a balance must be reached between the required time to get an analytical result and the expected level of accuracy. As an example, in a winery it is more important to sort different musts within a harvest according to their anthocyanin content than to know their absolute anthocyanin content.

The advent of ATR-MIR instruments has made it possible to obtain analytical results in a fast and easy manner, providing comparable sensitivities to commercial FTIR wine analyzers[207,208]. Alcohol content in several types of alcoholic beverages, including beer, wine and liquors, was successfully predicted using ATR-MIR and PLSR, achieving low RMSEP values for the validation sets[209]. The capability of ATR-MIR and PLSR to predict the concentration of total soluble solids (TSS, °Brix), pH, total phenolics, ammonia and free amino nitrogen was assessed in white grape juice samples, obtaining low standard errors of cross-validation (SECV) for all parameters[210]. The ferric reducing-antioxidant power assay (FRAP), the reference analysis for the quantification of the antioxidant capacity in wine, was correlated with ATR-MIR data in red wine samples, obtaining an RMSEP for the FRAP value of 4.7 mmol of Fe^{2+} per litre of wine[211].

1.5.3. Wine fermentation monitoring using MIR spectroscopy

As mentioned in section 1, wine fermentation is a complex process in which grape must is transformed into wine, involving the action of yeasts. In winemaking, monitoring of alcoholic fermentation is one of the most critical steps that defines if the final wine will be within the desired specifications. The wine industry needs fast and reliable process quality control methods that can operate in real time in order to assure the quality of the final product. Vibrational spectroscopy techniques have emerged as a powerful tool, not only for the analysis of must and wine samples, but also for process control and monitoring of the wine fermentation process.

Several authors have reported the use of ATR-MIR spectroscopy to monitor different parameters during alcoholic fermentation. Glucose, fructose, ethanol and glycerol were determined using ATR-MIR during red wine fermentation, with RMSECV values of <2 $g \cdot L^{-1}$ for glucose and fructose, 2.13 $g \cdot L^{-1}$ for ethanol and 0.42 $g \cdot L^{-1}$ for glycerol[212]. In a similar study, Buratti *et al.* monitored the progression of eight microvinifications using ATR-MIR and PLSR. They obtained high correlation coefficients when explaining the linear relationship between the kinetic parameters of chemical changes and the kinetic parameters obtained by MIR[213].

Using ATR-MIR spectra and PLSR for the determination of total antioxidant activity and total phenolic compounds, high correlations were found in red wines, while low correlations were found for rosé and white wines. This result suggests that, depending on the content of phenolic compounds, the accuracy of the models for the determination of antioxidant capacity will vary[214]. As it happened with the determination of organic acids, the quantification of antioxidant activity with ATR-MIR will vary depending on the concentration range, as the total antioxidant activity in rosé and white wines is low in comparison to red wines. A dispersive MIR spectrometer coupled with an ATR sensor was used to build PLSR models for on-line determination of glucose, fructose, ethanol, galactose, lactose and lactic acid in an industrial fermentation process. All SEP values were below 4 $g \cdot L^{-1}$, suggesting that ATR sensors can be used for alcoholic fermentation monitoring[104]. Fayolle *et al.* studied the effect of temperature in ATR-MIR spectra during alcoholic fermentation. They confirmed that the absorbance increased as the temperature decreased during the ATR-MIR analysis. but they obtained satisfactory SEP values in spite of temperature fluctuations: 5 $g \cdot L^{-1}$ for glucose, 5.8 $g \cdot L^{-1}$ for fructose, 0.6 $g \cdot L^{-1}$ for glycerol and 1.7 $g \cdot L^{-1}$ for ethanol[215]. The prediction of sulphate concentration was also investigated

in wines using FTIR spectroscopy and PLS-based calibration models (R= 0.98, LOD= 0.33 $g \cdot L^{-1}$), pointing out the suitability of this method to quantify this compound[216]. Fragoso *et al.* investigated the determination of total phenolic compounds ($mg \cdot L^{-1}$ gallic acid), total anthocyanins ($mg \cdot L^{-1}$ malvidin-3-glucoside), and condensed tannins ($mg \cdot L^{-1}$ catechin) during red wine fermentation using FT-IR in the transmission mode and PLSR. Their results suggest that FT-IR spectra coupled to PLSR are able to monitor the phenolic extraction during alcoholic fermentation[217]. Finally, red wine micro-vinifications were conducted to predict fermentation parameters using ATR-MIR and PLSR. The RMSECV values were 1.95 $g \cdot L^{-1}$ for glucose, 1.85 $g \cdot L^{-1}$ for fructose, 2.13 $g \cdot L^{-1}$ for ethanol and 0.42 g L^{-1} for glycerol[212].

All these studies confirm the suitability of MIR spectroscopy and chemometrics to control the wine fermentation process. A fast and accurate process monitoring is achieved not only for substrate conversion (e.g. sugars to ethanol) but also for the determination of product quality through the determination of some quality control parameters. Thus, PLSR models based on ATR-MIR data have emerged as rapid, easy and low-cost solutions to support or replace the commonly used reference analysis. The ability to provide fast quantitative chemical information for multiple compounds from a single measurement is also a strong advantage of using ATR-MIR spectroscopy and chemometrics[218].

1.5.4. The apparition of portable IR instruments and IR sensors

The appearance of hand-held and portable ATR-MIR spectrometers, or even ATR-MIR sensors for in- or on-line analysis, has allowed a much faster determination of chemical parameters in food and beverage products, and with good levels of accuracy. Several NIR and MIR hand-held and portable infrared spectrometers were used and compared to identify the presence of several adulterants in milk samples. Using PLSR regression, less than 1% of adulteration was detected for all the adulterants[219]. A portable ATR-MIR was used to build PLSR regression models to predict trans fats in edible oils, achieving low SEP values and estimating the limit of detection of the handheld instrument at 1% of trans by weight of fat[220]. A comparison between a hand-held and a benchtop ATR-MIR to monitor oil oxidative stability during a frying process showed similar RMSECV values for the PLSR models to predict the peroxide value, the acid value and the fatty acid composition[221]. Other applications of ATR-MIR portable spectrometers include the determination of alcohol mixtures or the quantitative determination of hydrocarbon contamination[222]. The determination of alcoholic strength, density and total dry extract in wine samples was also

achieved using an infrared sensor[223]. The use of portable ATR-MIR instruments and sensors together with chemometrics in the wine industry has led to the development of fully automated processes, making it possible to develop novel PAT tools to be applied in real-time during processing and manufacturing[224].

1.5.5. Detection of problems during alcoholic fermentation

Problems during wine alcoholic fermentation can result in economic loss and the production of lower quality wines. Hence, the improvement of the prediction capability for problematic fermentations would enable winemakers to take timely corrective actions and significantly reduce problematic fermentations. To this aim, preventive control may be possible, based on modelling of the wine fermentation process, as well as the integration of such models into control systems for optimising fermentation[225].

Hérnandez *et al.*, developed a method to detect abnormal wine fermentations using MIR spectral measurements during alcoholic fermentation and a multivariate classification method called Support Vector Machines (SVM). SVM correctly predicted 88% of the fermentation behaviour at 48 h from the beginning of fermentation[226]. Using an artificial must, Urtubia *et al.* studied two problematic fermentations: a fermentation with a temperature gradient and a nitrogen deficient fermentation. They found that ATR-MIR spectra and PLSR models for glycerol, ethanol and acetic acid allowed the discrimination of those problematic fermentations from the normal ones only after two days from inoculation time[227]. Applying different multivariate statistical and pattern recognition techniques, the behaviour of faulty wine fermentation batches was correctly predicted after 72 h or three days of fermentation[228].

The advantages of using IR spectroscopy coupled to chemometrics for process control monitoring are evident. These methodologies can serve as the basis for establishing PAT tools in the wine industry, initially in wine fermentation, but they could be transferred to other stages of wine production, such as grape ripening or wine aging.

1.6. References

1. International Organisation of Vine and Wine. *State of the World Vitivinicultural Sector in 2019.*; 2020. http://www.oiv.int/public/medias/7298/oiv-state-of-the-vitivinicultural-sector-in-2019.pdf.
2. Rodríguez Alonso P, Del Pozo de la Calle S, Valero Gaspar T, Ruiz Moreno E, Ávila Torres JM, Varela-Moreiras G. Fifty years of beverages consumption trends in Spanish households. *Nutr Hosp.* 2016;***33***:316. doi:10.20960/nh.316.

3. Varela-Moreiras G, Ávila JM, Cuadrado C, del Pozo S, Ruiz E, Moreiras O. Evaluation of food consumption and dietary patterns in Spain by the Food Consumption Survey: Updated information. *Eur J Clin Nutr.* 2010;**64**:S37-S43. doi:10.1038/ejcn.2010.208.

4. Iriarte-Chiapusso MJ, Ocete-Pérez CA, Hernández-Beloqui B, Ocete-Rubio R. Vitis vinifera in the Iberian Peninsula: A review. *Plant Biosyst.* 2017;**151**(2):245-257. doi:10.1080/11263504.2016.1165751.

5. Estreicher SK. A Brief History of Wine in Spain. *Eur Rev.* 2013;**21**(2):209-239. doi:10.1017/S1062798712000373.

6. This P, Lacombe T, Thomas MR. Historical origins and genetic diversity of wine grapes. *Trends Genet.* 2006;**22**(9):511-519. doi:10.1016/j.tig.2006.07.008.

7. De Andrés MT, Benito A, Pérez-Rivera G, *et al.* Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. *Mol Ecol.* 2012;**21**(4):800-816. doi:10.1111/j.1365-294X.2011.05395.x.

8. Keller M. *Botany and Anatomy.*; 2015. doi:10.1016/b978-0-12-419987-3.00001-7.

9. Myles S, Boyko AR, Owens CL, *et al.* Genetic structure and domestication history of the grape. *Proc Natl Acad Sci U S A.* 2011;**108**(9):3530-3535. doi:10.1073/pnas.1009363108.

10. Terral JF, Tabard E, Bouby L, *et al.* Evolution and history of grapevine (Vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann Bot.* 2010;**105**(3):443-455. doi:10.1093/aob/mcp298.

11. Ministerio de Agricultura y Pesca A y MA. Encuesta de viñedo 2015. *Secr Gen Técnica Subdirección Gen Estadística, Área Estadísticas Aliment.* 2015:1-68. http://www.mapama.gob.es/es/estadistica/temas/estadisticas-agrarias/memofinalvinedo_tcm7-443391.pdf.

12. Generalitat de Catalunya. Departament d'Agricultura, Ramaderia P i A, (INCAVI) IC de la V i del V. Comercialització any 2019. Denominacions d'Origen Protegides Catalanes. 2020:1-79.

13. Denominacions d'Origen catalanes. (INCAVI), Institut Català de la Vinya i el Vi. http://incavi.gencat.cat/ca/coneix-vi-catala/denominacions-origen-catalanes/. Accessed May 16, 2021.

14. Zoecklein BW, Fugelsang KC, Gump BH. *Practical Methods of Measuring Grape Quality.* Woodhead Publishing Limited; 2010. doi:10.1533/9781845699284.2.107.

15. Poni S, Gatti M, Palliotti A, *et al.* Grapevine quality: A multiple choice issue. *Sci Hortic (Amsterdam).* 2018;**234**(May 2017):445-462. doi:10.1016/j.scienta.2017.12.035.

16. Webb AD. Enology science and technology improve an ancient art. *Interdiscip Sci Rev.* 1980;**5**(1):49-59. doi:10.1179/030801880789767873.

17. Pretorius IS, Høj PB. Grape and wine biotechnology: Challenges, opportunities and potential benefits. *Aust J Grape Wine Res.* 2005;**11**(2):83-108. doi:10.1111/j.1755-0238.2005.tb00281.x.

18. Pretorius IS. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast.* 2000;**16**(8):675-729. doi:10.1002/1097-0061(20000615)16:8<675::aid-yea585>3.3.co;2-2.

19. Bisson LF. The biotechnology of wine yeast. *Food Biotechnol.* 2004;**18**(1):63-96. doi:10.1081/FBT-120030385.

20. Ribéreau-Gayon P, Dubourdieu D, Donèche B, Lonvaud A. *Handbook of Enology, Vol. 1: The Microbiology of Wine and Vinifications.* 2nd Editio. John Wiley & Sons; 2006. doi:10.1002/0470010363.fmatter.

21. Jackson RS. Chemical Constituents of Grapes and Wine. *Wine Sci.* 2000:232-280. doi:10.1016/b978-012379062-0/50007-x.

22. Zamora F. Biochemistry of Alcoholic Fermentation. In: Moreno-Arribas MV, Polo MC, eds. *Wine Chemistry and Biochemistry.* New York, NY: Springer New York; 2009:1-26. doi:10.1007/978-0-387-74118-5.

23. du Toit M, Pretorius IS. Microbial Spoilage and Preservation of Wine: Using Weapons from Nature's Own Arsenal -A Review. *South African J Enol Vitic.* 2019;**21**(1):74-96. doi:10.21548/21-1-3559.

24. Barata A, Malfeito-Ferreira M, Loureiro V. The microbial ecology of wine grape berries. *Int J Food Microbiol.* 2012;**153**(3):243-259. doi:10.1016/j.ijfoodmicro.2011.11.025.

25. Dussap CG, Poughon L. *Microbiology of Alcoholic Fermentation.* Elsevier B.V.; 2017. doi:10.1016/B978-0-444-63666-9.00010-8.

26. Tang K, Li Q. Biochemistry of Wine and Beer Fermentation. *Curr Dev Biotechnol Bioeng Food Beverages Ind.* 2017:281-304. doi:10.1016/B978-0-444-63666-9.00011-X.

27. Parapouli M, Vasileiadis A, Afendra AS, Hatziloukas E. *Saccharomyces Cerevisiae and Its Industrial Applications.* Vol 6.; 2020. doi:10.3934/microbiol.2020001.

28. Suárez-Lepe JA, Morata A. New trends in yeast selection for winemaking. *Trends Food Sci Technol.* 2012;**23**:39-50. doi:10.1016/j.tifs.2011.08.005.

29. Fleet GH. Growth of Yeasts during Wine Fermentations. *J Wine Res.* 1990;**1**(3):211-223. doi:10.1080/09571269008717877.

30. Malfeito-Ferreira M. *Wines: Wine Spoilage Yeasts and Bacteria.* Vol 3. Second Edi. Elsevier; 2014. doi:10.1016/B978-0-12-384730-0.00390-6.

31. Seo SO, Park SK, Jung SC, Ryu CM, Kim JS. Anti-contamination strategies for yeast fermentations. *Microorganisms.* 2020;**8**(2). doi:10.3390/microorganisms8020274.

32. Lonvaud-Funel A. Chapter 1: Undesirable Compounds and Spoilage Microorganisms in Wine. In: *Wine Safety, Consumer Preference, and Human Health.* Switzerland: Springer International Publishing; 2016:1-26. doi:10.1007/978-3-319-24514-0_1.

33. Osborne JP. *Advances in Microbiological Quality Control.* Woodhead Publishing Limited; 2010. doi:10.1533/9781845699284.2.162.

34. F.F. Bauer, I.S. Pretorius. Yeast Stress Response and Fermentation Efficiency: How to Survive the Making of Wine - A Review. *South African J Enol Vitic.* 2000;**21**(1):27-51. https://dialnet.unirioja.es/servlet/articulo?codigo=203548.

35. Alexandre H, Charpentier C. Biochemical aspects of stuck and sluggish fermentation in grape must. *J Ind Microbiol Biotechnol.* 1998;**20**:20-27. doi:10.1038/sj.jim.2900442.

36. Bisson LF, Butzke CE. Diagnosis and rectification of stuck and sluggish fermentations. *Am J Enol Vitic.* 2000.

37. Bisson LF. Stuck and sluggish fermentations. *Am J Enol Vitic.* 1999;**50**(1):107-119.

38. Salmon JM, Barre P. Improvement of nitrogen assimilation and fermentation kinetics under enological conditions by derepression of alternative nitrogen-assimilatory pathways in an industrial *Saccharomyces cerevisiae* strain. *Appl Environ Microbiol.* 1998;**64**(10):3831-3837. doi:10.1128/aem.64.10.3831-3837.1998.

39. Gobert A, Tourdot-Maréchal R, Sparrow C, Morge C, Alexandre H. Influence of nitrogen status in wine alcoholic fermentation. *Food Microbiol.* 2019;**83**(November 2018):71-85. doi:10.1016/j.fm.2019.04.008.

40. Martínez-Moreno R, Quirós M, Morales P, Gonzalez R. New insights into the advantages of ammonium as a winemaking nutrient. *Int J Food Microbiol.* 2014;**177**:128-135. doi:10.1016/j.ijfoodmicro.2014.02.020.

41. Delfini C, Formica J V. Chapter 16: Microbiological monitoring of must and wine. In: *Wine Microbiology: Science and Technology.* Taylor & Francis Group; 2001.

42. Esteve-Zarzoso B, Martínez M, Rubires X, Yuste-Rojas M, Torres M. Applied Wine Microbiology. In: *Molecular Wine Microbiology.* ; 2011:341-355. doi:10.1016/B978-0-12-375021-1.10014-1.

43. Lonvaud-Funel A. Effects of malolactic fermentation on wine quality. In: *Managing Wine Quality.* Elsevier; 2010:60-92. doi:10.1533/9781845699987.1.60.

44. Muñoz R, Moreno-Arribas MV, Rivas B de las. Lactic Acid Bacteria. In: *Molecular Wine Microbiology.* ; 2011. doi:10.1016/B978-0-12-375021-1.10008-6.

45. Gil-Sánchez I, Bartolomé Suáldea B, Victoria Moreno-Arribas M. Malolactic Fermentation. In: *Red Wine Technology.* Elsevier; 2019:85-98. doi:10.1016/B978-0-12-814399-5.00006-2.

46. Ferreira AM, Mendes-Faia A. The role of yeasts and lactic acid bacteria on the metabolism of organic acids during winemaking. *Foods.* 2019;**9**(9). doi:10.3390/foods9091231.

47. Fugelsang KC, Edwards CG. Lactic Acid Bacteria. In: Fugelsang KC, Edwards CG, eds. *Wine Microbiology: Practical Applications and Procedures.* Boston, MA: Springer US; 2007:1-393. doi:10.1007/978-0-387-33349-6.

48. Du Plessis HW, Steger CLC, Du Toit M, Lambrechts MG. The occurrence of malolactic fermentation in brandy base wine and its influence on brandy quality. *J Appl Microbiol.* 2002;**92**(5):1005-1013. doi:10.1046/j.1365-2672.2002.01616.x.

49. Lafon-Lafourcade S, Lonvaud-Funel A, Carre E. Lactic acid bacteria of wines: stimulation of growth and malolactic fermentation. *Antonie Van Leeuwenhoek.* 1983;**49**(3):349-352. doi:10.1007/BF00399509.

50. Versari A, Patrizi C, Parpinello GP, *et al.* Effect of co-inoculation with yeast and bacteria on chemical and sensory characteristics of commercial Cabernet Franc red wine from Switzerland. *J Chem Technol Biotechnol.* 2016;**91**(4):876-882. doi:10.1002/jctb.4652.

51. Gomes RJ, Borges M de F, Rosa M de F, Castro-Gómez RJH, Spinosa WA. Acetic acid bacteria in the food industry: Systematics, characteristics and applications. *Food Technol Biotechnol.* 2018;**56**(2):139-151. doi:10.17113/ftb.56.02.18.5593.

52. Guillamón JM, Mas A. Acetic Acid Bacteria. *Mol Wine Microbiol.* 2011:227-255. doi:10.1016/B978-0-12-375021-1.10009-8.

53. Saichana N, Matsushita K, Adachi O, Frébort I, Frebortova J. Acetic acid bacteria: A group of bacteria with versatile biotechnological applications. *Biotechnol Adv*. 2015;**33**(6):1260-1271. doi:10.1016/j.biotechadv.2014.12.001.

54. Bartowsky EJ, Xia D, Gibson RL, Fleet GH, Henschke PA. Spoilage of bottled red wine by acetic acid bacteria. *Lett Appl Microbiol*. 2003;**36**(5):307-314. doi:10.1046/j.1472-765X.2003.01314.x.

55. Lesschaeve I. Sensory Evaluation of Wine and Commercial Realities: Review of Current Practices and Perspectives. *Am J Enol Vitic*. 2007;**58**(2):252-258. http://www.ajevonline.org/content/ajev/58/2/252.full.pdf.

56. Langstaff SA. *Sensory Quality Control in the Wine Industry*. Woodhead Publishing Limited; 2010. doi:10.1533/9781845699512.3.236.

57. OIV International. *Review Document on Sensory Analysis of Wine*.; 2015. http://www.oiv.int/public/medias/3307/review-on-sensory-analysis-of-wine.pdf.

58. Lesschaeve I, Noble AC. *Sensory Analysis of Wine*. Woodhead Publishing Limited; 2010. doi:10.1533/9781845699284.2.189.

59. International Organisation of Vine and Wine. *Compendium of International Methods of Wine and Must Analysis*.; 2012.

60. 9/2000 RO. Compedium of International Methods of Analysis OIV Classification. 2009:2.

61. European Union. *Commission Regulation (EC) No 606/2009*.; 2009:1-5. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2006R1881:20100701:EN:PDF%0Ahttps://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01985L0374-19990604&qid=1604918047856.

62. Pramod K, Tahir Ma, Charoo N, Ansari S, Ali J. Pharmaceutical product development: A quality by design approach. *Int J Pharm Investig*. 2016;**6**(3):129. doi:10.4103/2230-973x.187350.

63. Rathore AS, Kapoor G. Implementation of quality by design toward processing of food products. *Prep Biochem Biotechnol*. 2017;**47**(5):435-440. doi:10.1080/10826068.2017.1315601.

64. Rathore AS. QbD/PAT for bioprocessing: Moving from theory to implementation. *Curr Opin Chem Eng*. 2014;**6**:1-8. doi:10.1016/j.coche.2014.05.006.

65. Food and Drug Administration. Guidance for Industry. PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. 2004;(September). http://www.fda.gov/cvm/guidance/published.html. Accessed September 2, 2020.

66. Lopes JA, Costa PF, Alves TP, Menezes JC. Chemometrics in bioprocess engineering: Process analytical technology (PAT) applications. In: *Chemometrics and Intelligent Laboratory Systems*. ; 2004. doi:10.1016/j.chemolab.2004.07.006.

67. Van Den Berg F, Lyndgaard CB, Sørensen KM, Engelsen SB. Process Analytical Technology in the food industry *. *Trends Food Sci Technol*. 2013;**31**:27-35. doi:10.1016/j.tifs.2012.04.007.

68. Chew W, Sharratt P. Trends in process analytical technology. *Anal Methods*. 2010;**2**(10):1412-1438. doi:10.1039/c0ay00257g.

69. Yao S, Song H, Cheng Q, Liang B. The Development of PAT (Process Analytical Technology) for Drug Production and the Requirements for Domestic Pharmaceutical Engineering Education. *Creat Educ.* 2012;**03**(07):76-79. doi:10.4236/ce.2012.37b019.

70. Marison I, Hennessy S, Foley R, Schuler M, Sivaprakasam S, Freeland B. The choice of suitable online analytical techniques and data processing for monitoring of bioprocesses. *Adv Biochem Eng Biotechnol.* 2013;**132**:249-280. doi:10.1007/10_2012_175.

71. Roychoudhury P, Harvey LM, McNeil B. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy. *Anal Chim Acta.* 2006;**561**(1-2):218-224. doi:10.1016/j.aca.2006.01.037.

72. Claßen J, Aupert F, Reardon KF, Solle D, Scheper T. Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. *Anal Bioanal Chem.* 2017;**409**(3):651-666. doi:10.1007/s00216-016-0068-x.

73. Faassen SM, Hitzmann B. Fluorescence spectroscopy and chemometric modeling for bioprocess monitoring. *Sensors (Switzerland).* 2015;**15**(5):10271-10291. doi:10.3390/s150510271.

74. Lee JD, Seppelt BD. Human Factors in Automation Design. *Springer Handb Autom.* 2009:417-436. doi:10.1007/978-3-540-78831-7_25.

75. Kueppers S, Haider M. Process analytical chemistry-future trends in industry. *Anal Bioanal Chem.* 2003;**376**(3):313-315. doi:10.1007/s00216-003-1907-0.

76. Sorak D, Herberholz L, Iwascek S, Altinpinar S, Pfeifer F, Siesler HW. New developments and applications of handheld raman, mid-infrared, and near-infrared spectrometers. *Appl Spectrosc Rev.* 2012;**47**(2):83-115. doi:10.1080/05704928.2011.625748.

77. Chen W, Yao Y, Chen T, Shen W, Tang S, Lee HK. Application of smartphone-based spectroscopy to biosample analysis: A review. *Biosens Bioelectron.* 2021;**172**. doi:10.1016/j.bios.2020.112788.

78. Cullen PJ, O'Donnell CP, Fagan CC. Benefits and Challenges of Adopting PAT for the Food Industry. In: *Food Engineering Series.* ; 2014:1-5. doi:10.1007/978-1-4939-0311-5_1.

79. Fellows PJ. Properties of food and principles of processing. In: *Food Processing Technology.* ; 2017:3-200. doi:10.1016/b978-0-08-100522-4.00001-8.

80. Varzakas T. ISO 22000, HACCP, and Other Management Tools for Implementation of Food Safety-Traceability Case Studies. In: *Title: Handbook of Food Processing : Food Safety, Quality, and Manufacturing Processes.* Boca raton, Florida: CRC Press; 2016.

81. Nielsen SS. Introduction to Food Analysis. In: *Food Analysis.* Springer International Publishing; 2017:3-16. doi:10.1007/978-3-319-45776-5_1.

82. Mohamed GF, Shaheen MS, Khalil SKH, Hussein AMS, Kamil MM. Application of FT-IR Spectroscopy for Rapid and Simultaneous Quality Determination of Some Fruit Products. *Nat Sci.* 2011;**9**(11):21-31. http://www.sciencepub.net/nature.

83. Sørensen KM, Khakimov B, Engelsen SB. The use of rapid spectroscopic screening methods to detect adulteration of food raw materials and ingredients. *Curr Opin Food Sci.* 2016;**10**:45-51. doi:10.1016/j.cofs.2016.08.001.

84. Porep JU, Kammerer DR, Carle R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci Technol.* 2015;**46**(2):211-230. doi:10.1016/j.tifs.2015.10.002.

85. Fukuda IM, Pinto CFF, Moreira CDS, Saviano AM, Lourenço FR. Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). *Brazilian J Pharm Sci.* 2018;**54**(Special Issue). doi:10.1590/s2175-97902018000001006.

86. Zhao M, Esquerre C, Downey G, O'Donnell CP. Process analytical technologies for fat and moisture determination in ground beef - a comparison of guided microwave spectroscopy and near infrared hyperspectral imaging. *Food Control.* 2017;**73**:1082-1094. doi:10.1016/j.foodcont.2016.10.023.

87. Tajammal Munir M, Yu W, Young BR, Wilson DI. The current status of process analytical technologies in the dairy industry. *Trends Food Sci Technol.* 2015;**43**(2):205-218. doi:10.1016/j.tifs.2015.02.010.

88. Panikuttira B, O'Shea N, Tobin JT, Tiwari BK, O'Donnell CP. Process analytical technology for cheese manufacture. *Int J Food Sci Technol.* 2018;**53**(8):1803-1815. doi:10.1111/ijfs.13806.

89. El-Abassy RM, Donfack P, Materny A. Rapid determination of free fatty acid in extra virgin olive oil by raman spectroscopy and multivariate analysis. *JAOCS, J Am Oil Chem Soc.* 2009;**86**(6):507-511. doi:10.1007/s11746-009-1389-0.

90. Jerome RE, Singh SK. Process analytical technology for bakery industry : A review. 2019;(March):1-21. doi:10.1111/jfpe.13143.

91. Xiaping F, Yibin Y, Ying Z, Huishan L, Huirong X. Quantitative and qualitative measurement of pear firmness based on near infrared spectroscopy and chemometrics. *Int J Agric Biol Eng.* 2008;**1**(1):69-74. doi:10.3965/j.issn.1934-6344.2008.01.069-074.

92. Mendoza F, Lu R, Cen H. Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. *Postharvest Biol Technol.* 2012;**73**:89-98. doi:10.1016/j.postharvbio.2012.05.012.

93. OIV. "Guidelines on Infrared Analysers in Oenology." *Resolut OIV/Oeno 390/2010.* 2010;(June):1-8. http://www.oiv.int/public/medias/379/viti-2010-1-en.pdf.

94. Menezes JC. Process Analytical Technology in Bioprocess Development and Manufacturing. *Compr Biotechnol.* 2011;**3**:501-509. doi: 10.1016/B978-0-08-088504-9.00205-1

95. Caplice E, Fitzgerald GF. Food fermentations: Role of microorganisms in food production and preservation. *Int J Food Microbiol.* 1999;**50**(1-2):131-149. doi:10.1016/S0168-1605(99)00082-3.

96. Landgrebe D, Haake C, Höpfner T, *et al.* On-line infrared spectroscopy for bioprocess monitoring. *Appl Microbiol Biotechnol.* 2010;**88**(1):11-22. doi:10.1007/s00253-010-2743-8.

97. Scheper T, Hitzmann B, Stärk E, *et al.* Bioanalytics: Detailed insight into bioprocesses. *Anal Chim Acta.* 1999;**400**(1-3):121-134. doi:10.1016/S0003-2670(99)00612-1.

98. Mello LD, Kubota LT. Review of the use of biosensors as analytical tools in the food and drink industries. *Food Chem.* 2002;**77**(2):237-256. doi:10.1016/S0308-8146(02)00104-8.

99. Luong JHT, Bouvrette P, Male KB. Developments and applications of biosensors in food analysis. *Trends Biotechnol.* 1997;**15**(September):369-377.

100. Vojinović V, Cabral JMS, Fonseca LP. Real-time bioprocess monitoring: Part I: In situ sensors. *Sensors Actuators, B Chem.* 2006;**114**(2):1083-1091. doi:10.1016/j.snb.2005.07.059.

101. Colombié S, Malherbe S, Sablayrolles JM. Modeling alcoholic fermentation in enological conditions: Feasibility and interest. *Am J Enol Vitic.* 2005.

102. Schalk R, Geoerg D, Staubach J, Raedle M, Methner F-J, Beuermann T. Evaluation of a newly developed mid-infrared sensor for real-time monitoring of yeast fermentations. *J Biosci Bioeng.* 2017;**123**:651-657. doi:10.1016/j.jbiosc.2016.12.005.

103. Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B. On-line fermentation monitoring by mid-infrared spectroscopy. *Appl Spectrosc.* 2004;**58**(7):804-810. doi:10.1366/0003702041389229.

104. Fayolle P, Picque D, Corrieu G. On-line monitoring of fermentation processes by a new remote dispersive middle-infrared spectrometer. *Food Control.* 2000;**11**(4):291-296. doi:10.1016/S0956-7135(99)00105-X.

105. Ündey C, Ertun S, Mistretta T, Looze B. Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control. In: *Journal of Process Control.* Vol 20. ; 2010:1009-1018. doi:10.1016/j.jprocont.2010.05.008.

106. Sun D-W. *Infrared Spectroscopy for Food Quality Analysis and Control.* Amsterdam : Academic Press; 2009.

107. Penner MH. Basic Principles of Spectroscopy. In: *Food Analysis.* Springer International Publishing; 2017:79-88. doi:10.1007/978-3-319-45776-5_6.

108. Larkin P. Basic Principles. In: *Infrared and Raman Spectroscopy.* Elsevier; 2011:7-25. doi:10.1016/B978-0-12-386984-5.10002-3.

109. Odularu AT. Worthwhile Relevance of Infrared Spectroscopy in Characterization of Samples and Concept of Infrared Spectroscopy-Based Synchrotron Radiation. *J Spectrosc.* 2020;**2020**:1-11. doi:10.1155/2020/8869713.

110. Sun D-W. *Infrared Spectroscopy for Food Quality Analysis and Control.*; 2009. doi:10.1016/B978-0-12-374136-3.00010-9.

111. Stuart BH. *Infrared Spectroscopy: Fundamentals and Applications.* Vol 8.; 2004. doi:10.1002/0470011149.

112. Bauer R, Bauer FF, Kossmann J, Koch KR, Esbensen KH. FTIR Spectroscopy for Grape and Wine Analysis. *Anal Chem.* 2008;**80**(5):1371-1379. https://pubs.acs.org/sharingguidelines. Accessed July 26, 2018.

113. Sun DW. *Infrared Spectroscopy for Food Quality Analysis and Control.* (Edition F, ed.). Academic Press; 2009. doi:10.1016/B978-0-12-374136-3.X0001-6.

114. Rodriguez-Saona L, Ayvaz H, Wehling RL. Infrared and Raman Spectroscopy. In: *Handbook of Grignard Reagents.* CRC Press; 2017:107-127. doi:10.1007/978-3-319-45776-5_8.

115. Abbas O, Pissard A, Baeten V. *Near-Infrared, Mid-Infrared, and Raman Spectroscopy.*

Second Edi. Elsevier Inc.; 2020. doi:10.1016/b978-0-12-813266-1.00003-6.

116. Stuart BH. *Infrared Spectroscopy: Fundamentals and Applications.*; 2005. doi:10.1002/0470011149.

117. Griffiths PR, De Haseth JA. *Fourier Transform Infrared Spectrometry.* Wiley-Interscience; 2007.

118. Ganzoury MA, Allam NK, Nicolet T, All C. Introduction to Fourier Transform Infrared Spectrometry. *Thermo Nicolet.*:1-8. doi:10.1016/j.rser.2015.05.073.

119. Coates JP. *Infrared Spectroscopy for Process Analytical Applications.*; 2010. doi:10.1002/9780470689592.ch6.

120. Hashimoto A, Kameoka T. Applications of infrared spectroscopy to biochemical, food, and agricultural processes. *Appl Spectrosc Rev.* 2008;**43**(5):416-451. doi:10.1080/05704920802108131.

121. Smith BC. Fundamentals of Molecular Absorption Spectroscopy. *Quant Spectrosc Theory Pract.* 2002:1-41. doi:10.1016/b978-012650358-6/50002-6.

122. Craig AP, Franca AS, Irudayaraj J. Vibrational spectroscopy for food quality and safety screening. In: *High Throughput Screening for Food Safety Assessment: Biosensor Technologies, Hyperspectral Imaging and Practical Applications.* 1st Editio. Elsevier Ltd; 2015:165-194. doi:10.1016/B978-0-85709-801-6.00007-1.

123. Blanco M, Villarroya I. NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends Anal Chem.* 2002;**21**(4):240-250. doi:10.1016/S0165-9936(02)00404-1.

124. Griffiths PR. Introduction to the Theory and Instrumentation for Vibrational Spectroscopy. 2006. doi:10.1002/9780470027325.s8935.

125. Huang H, Yu H, Xu H, Ying Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J Food Eng.* 2008;**87**(3):303-313. doi:10.1016/j.jfoodeng.2007.12.022.

126. O'Donnell C, Fagan C, Cullen PJ. *Process Analytial Technology for the Food Industry.*; 2009. doi:10.1016/S1571-5078(08)00422-4.

127. Daham B, Andrews GE, Li H, *et al.* Application of a portable FTIR for measuring on-road emissions. *SAE technical paper.* 2005-01-0676. doi:10.4271/2005-01-0676.

128. Higgins F. *Rapid and Reliable Phthalate Screening in Plastics by Portable FTIR Spectroscopy Application Note Consumer Products.* https://www.cpsc.gov/phthalates.

129. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory--Part 1: Spectroscopic techniques. *Forensic Sci Int.* 2014;**245**:187-194. doi:10.1016/j.forsciint.2014.08.009.

130. Underhill SV and M. Raman and infrared techniques for fighting drug-related crime: a preliminary assessment. *Proc SPIE 6402, Opt Photonics Counterterrorism Crime Fight II.* 2006. doi:10.1117/12.689099.

131. Ayvaz H, Rodriguez-Saona LE. Application of handheld and portable spectrometers for screening acrylamide content in commercial potato chips. *Food Chem.* 2015;**174**:154-162. doi:10.1016/j.foodchem.2014.11.001.

132. Mossoba MM, Srigley CT, Kramer JKG. Portable FTIR analyzers for the rapid determination of total trans fat. *Lipid Technol.* 2015;**27**(1):11-13.

doi:10.1002/lite.201400074.

133. Spragg RA. IR Spectroscopy Sample Preparation Methods in *Encyclopedia of Spectroscopy and Spectrometry*. Elsevier. 1999; 1058-1066..

134. Fagan CC, O'Donnell CP. Application of Mid-Infrared Spectroscopy to Food Processing Systems. *Nondestruct Test Food Qual.* 2008;(Reh 2001):119-142. doi:10.1002/9780470388310.ch5.

135. Höpe A. Diffuse Reflectance and Transmittance. *Experimental Methods in the Physical Sciences* . 2014; **46**: 179-219. doi:10.1016/B978-0-12-386022-4.00006-6.

136. Coates J. A review of sampling methods for infrared spectroscopy. In: Workman J, Springsteen AW, eds. *Applied Spectroscopy: A Compact Reference for Practitioners*. New York: Academic Press; 1998:49-91. doi:10.1016/B978-012764070-9/50005-6.

137. Miller CE. Chemometrics in Process Analytical Technology (PAT) in*Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*. Second Edition; 2010; 353-438 doi:10.1002/9780470689592.ch12.

138. Mercier SM, Diepenbroek B, Wijffels RH, Streefland M. Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations. *Trends Biotechnol.* 2014;**32**(6):329-336. doi:10.1016/j.tibtech.2014.03.008.

139. Figueiredo Sales R, Vitale R, Moreira de Lima S, Pimentel MF, Stragevitch L, Ferrer A. Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy. *Comput Chem Eng.* 2016;**94**:343-353. doi:10.1016/j.compchemeng.2016.08.013.

140. Manley M, van Zyl A, Wolf EEH. The Evaluation of the Applicability of Fourier Transform Near-Infrared (FT-NIR) Spectroscopy in the Measurement of Analytical Parameters in Must and Wine. *South African J Enol Vitic.* 2001;**22**(2):93-100. doi:https://doi.org/10.21548/22-2-2201.

141. Engel J, Gerretzen J, Szymańska E, *et al.* Breaking with trends in pre-processing? *TrAC - Trends Anal Chem.* 2013;**50**:96-106. doi:10.1016/j.trac.2013.04.015.

142. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem.* 1964;**36**(8):1627-1639. doi:10.1021/ac60214a047.

143. Chuen Lee L, Liong C-Y, Aziz Jemain A. A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemom Intell Lab Syst.* 2017;**163**:64-75. doi:10.1016/j.chemolab.2017.02.008.

144. Oliveri P, Malegori C, Simonetti R, Casale M. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Anal Chim Acta.* 2019;**1058**:9-17. doi:10.1016/j.aca.2018.10.055.

145. Rabatel G, Marini F, Walczak B, Roger JM. VSN: Variable sorting for normalization. *J Chemom.* 2020;**34**(2). doi:10.1002/cem.3164.

146. Rinnan A, Van Den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal Chem.* 2009;**28**:1201-1222. doi:10.1016/j.trac.2009.07.007.

147. Andersson CA. Direct orthogonalization. *Chemom Intell Lab Syst.* 1999;**47**(1):51-63.

doi:10.1016/S0169-7439(98)00158-0.

148. Feudale RN, Tan H, Brown SD. *Piecewise Orthogonal Signal Correction*. www.eigenvector.com. Accessed October 24, 2019.

149. Roger JM, Chauchard F, Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom Intell Lab Syst*. 2003;**66**(2):191-204. doi:10.1016/S0169-7439(03)00051-0.

150. Liland KH. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends Anal Chem*. 2011;**30**(6):827-841. doi:10.1016/j.trac.2011.02.007.

151. Bro R, Smilde AK. Principal component analysis. *Anal Methods*. 2014;**6**(9):2812-2831. doi:10.1039/C3AY41907J.

152. Pasquini C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Anal Chim Acta*. 2018;**1026**:8-36. doi:10.1016/j.aca.2018.04.004.

153. Yun YH, Li HD, Deng BC, Cao DS. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC - Trends Anal Chem*. 2019;**113**:102-115. doi:10.1016/j.trac.2019.01.018.

154. Gerretzen J, Szymańska E, Bart J, *et al*. Boosting model performance and interpretation by entangling preprocessing selection and variable selection. *Anal Chim Acta*. 2016;**938**:44-52. doi:10.1016/j.aca.2016.08.022.

155. Cocchi M, Biancolillo A, Marini F. *Chemometric Methods for Classification and Feature Selection*. Vol 82. 1st ed. Elsevier B.V.; 2018. doi:10.1016/bs.coac.2018.08.006.

156. Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng*. 2014;**40**(1):16-28. doi:10.1016/j.compeleceng.2013.11.024.

157. Monakhova YB, Mushtakova SP. Methodology of Chemometric Modeling of Spectrometric Signals in the Analysis of Complex Samples. *J Anal Chem*. 2017;**72**(2):119-128. doi:10.1134/S1061934816120066.

158. Wold SJ. Principal components analysis. *Chemom Intell Lab Syst*. 1987;**2**:37-52. doi:10.1016/B978-0-08-044894-7.01358-0.

159. Geladi P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochim Acta - Part B At Spectrosc*. 2003;**58**(5):767-782. doi:10.1016/S0584-8547(03)00037-5.

160. Bro R, Smilde AK. Principal component analysis. *Anal Methods*. 2014;**6**(9):2812-2831. doi:10.1039/c3ay41907j.

161. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W, Koch RS. *PLS_Toolbox Version 4.0 for Use with MATLAB ™*.; 2006. http://mitr.p.lodz.pl/raman/jsurmacki/pliki/zajecia/LMDiT/cw4i5/LMDiT_PLS_Manual_4.pdf.

162. Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics*. 1979;**21**(3):341-349. doi:10.1080/00401706.1979.10489779.

163. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers RJAN, van der Greef J, Timmerman ME.

ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005;**21**(13):3043-3048. doi:10.1093/bioinformatics/bti476.

164. Jansen JJ, Hoefsloot HCJ, Van Der Greef J, Timmerman ME, Westerhuis JA, Smilde AK. ASCA: Analysis of multivariate data obtained from an experimental design. *J Chemom*. 2005;**19**(9):469-481. doi:10.1002/cem.952.

165. De Luca S, De Filippis M, Bucci R, Magrì AD, Magrì AL, Marini F. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. *Microchem J*. 2016;**129**:348-361. doi:10.1016/j.microc.2016.07.021.

166. Ryckewaert M, Gorretta N, Henriot F, Marini F, Roger JM. Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Anal Chim Acta*. 2020;**1101**:23-31. doi:10.1016/j.aca.2019.12.024.

167. Anderson MJ, Ter Braak CJF. Permutation tests for multi-factorial analysis of variance. *J Stat Comput Simul*. 2003;**73**(2):85-113. doi:10.1080/00949650215733.

168. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers RJAN, van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005;**21**(13):3043-3048. doi:10.1093/bioinformatics/bti476.

169. Smilde AK, Hoefsloot HCJ, Westerhuis JA. The geometry of ASCA. *J Chemom*. 2008;**22**(8):464-471. doi:10.1002/cem.1175.

170. Geladi P, Kowalski BR. Partial Least-squares Regression: a tutorial. *Anal Chim Acta*. 1986;**185**:1-17. doi:10.1016/0003-2670(86)80028-9.

171. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst*. 2001;**58**(2):109-130. doi:10.1016/S0169-7439(01)00155-1.

172. Kumar N, Bansal A, Sarma GS, Rawal RK. Chemometrics tools used in analytical chemistry: An overview. *Talanta*. 2014;**123**:186-199. doi:10.1016/j.talanta.2014.02.003.

173. Ballabio D, Consonni V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal Methods*. 2013;**5**(16):3790-3798. doi:10.1039/c3ay40582f.

174. Barker M, Rayens W. Partial least squares for discrimination. *J Chemom*. 2003;**17**(3):166-173. doi:10.1002/cem.785.

175. Pomerantsev AL, Rodionova OY. Multiclass partial least squares discriminant analysis: Taking the right way—A critical tutorial. *J Chemom*. 2018;**32**(8). doi:10.1002/cem.3030.

176. Pérez NF, Ferré J, Boqué R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom Intell Lab Syst*. 2009;**95**(2):122-128. doi:10.1016/j.chemolab.2008.09.005.

177. de Juan A, Tauler R. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. *Crit Rev Anal Chem*. 2006;**36**(3-4):163-176. doi:10.1080/10408340600970005.

178. De Juan A, Tauler R. Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution. In: *Analytica Chimica Acta*.

Vol 500. Elsevier; 2003:195–210. doi:10.1016/S0003-2670(03)00724-4.

179. de Juan A, Casassas E, Tauler R. Soft Modeling of Analytical Data. In: *Encyclopedia of Analytical Chemistry*. ; 2000. doi:10.1002/9780470027318.a5208.

180. Garrido M, Rius FX, Larrechi MS. Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Anal Bioanal Chem*. 2008;**390**(8):2059–2066. doi:10.1007/s00216-008-1955-6.

181. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: New features and applications. *Chemom Intell Lab Syst*. 2015;**140**:1–12. doi:10.1016/j.chemolab.2014.10.003.

182. Jaumot J, Vives M, Gargallo R. Application of multivariate resolution methods to the study of biochemical and biophysical processes. *Anal Biochem*. 2004. doi:10.1016/j.ab.2003.12.028.

183. Vier D, Wambach S, Schünemann V, Gollmer KU. Multivariate curve resolution and carbon balance constraint to unravel FTIR spectra from fed-batch fermentation samples. *Bioengineering*. 2017;**4**(1). doi:10.3390/bioengineering4010009.

184. Blanco M, Peinado AC, Mas J. Monitoring alcoholic fermentation by joint use of soft and hard modelling methods. *Anal Chim Acta*. 2006;**556**(2):364–373. doi:10.1016/j.aca.2005.09.066.

185. Ferrer A. Latent structures-based multivariate statistical process control: A paradigm shift. *Qual Eng*. 2014;**26**(1):72–91. doi:10.1080/08982112.2013.846093.

186. Albert S, Kinley RD. Multivariate statistical monitoring of batch processes: An industrial case study of fermentation supervision. *Trends Biotechnol*. 2001;**19**(2):53–62. doi:10.1016/S0167-7799(00)01528-6.

187. Rogalewicz M. Some Notes on Multivariate Statistical Process Control. *Manag Prod Eng Rev*. 2012;**3**(4):80–86. doi:10.2478/v10270-012-0036-7.

188. Kourti T. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *J Chemom*. 2003;**17**(1):93–109. doi:10.1002/cem.778.

189. Camacho J, Pérez-Villegas A, Garciá-Teodoro P, MacIá-Fernández G. PCA-based multivariate statistical network monitoring for anomaly detection. *Comput Secur*. 2016. doi:10.1016/j.cose.2016.02.008.

190. Kourti T. Process analytical technology beyond real-time analyzers: The role of multivariate analysis. *Crit Rev Anal Chem*. 2006;**36**(3–4):257–278. doi:10.1080/10408340600969957.

191. Lennox B, Montague GA, Hiden HG, Kornfeld G, Goulding PR. Process monitoring of an industrial fed-batch fermentation. *Biotechnol Bioeng*. 2001;**74**(2):125–135. doi:10.1002/bit.1102.

192. AlGhazzawi A, Lennox B. Monitoring a complex refining process using multivariate statistics. *Control Eng Pract*. 2008;**16**(3):294–307. doi:10.1016/j.conengprac.2007.04.014.

193. Catelani TA, Santos JR, Páscoa RNMJ, Pezza L, Pezza HR, Lopes JA. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta*. 2018;**179**(October 2017):292–299. doi:10.1016/j.talanta.2017.11.010.

194. Chapman J, Gangadoo S, Truong VK, Cozzolino D. Spectroscopic approaches for rapid beer and wine analysis. *Curr Opin Food Sci.* 2019;**28**:67-73. doi:10.1016/j.cofs.2019.09.001.

195. Ricci A, Olejar KJ, Parpinello GP, Kilmartin PA, Versari A. Application of Fourier transform infrared (FTIR) spectroscopy in the characterization of tannins. *Appl Spectrosc Rev.* 2015;**50**(5):407-442. doi:10.1080/05704928.2014.1000461.

196. Pearce K, Culbert J, Cass D, Cozzolino D, Wilkinson K. Influence of Sample Storage on the Composition of Carbonated Beverages by MIR Spectroscopy. *Beverages.* 2016;**2**(4):26. doi:10.3390/beverages2040026.

197. Cozzolino D, Curtin C. The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments. *Food Control.* 2012;**26**:241-246. doi:10.1016/j.foodcont.2012.02.006.

198. Wynne L, Clark S, Adams MJ, Barnett NW. Compositional dynamics of a commercial wine fermentation using two-dimensional FTIR correlation analysis. *Vib Spectrosc.* 2007;**44**(2):394-400. doi:10.1016/j.vibspec.2007.03.010.

199. Moreira JL, Santos L. Analysis of organic acids in wines by Fourier-transform infrared spectroscopy. In: *Analytical and Bioanalytical Chemistry.* ; 2005. doi:10.1007/s00216-005-3062-2.

200. Cozzolino D, Dambergs RG. Wine and Beer. In: *Infrared Spectroscopy for Food Quality Analysis and Control.* ; 2009. doi:10.1016/B978-0-12-374136-3.00014-6.

201. Patz C-D, Blieke A, Ristow R, Dietrich H. Application of FT-MIR spectrometry in wine analysis. *Anal Chim Acta.* 2004;**513**:81-89. doi:10.1016/j.aca.2004.02.051.

202. Kupina SA, Shrikhande AJ. Evaluation of a Fourier transform infrared instrument for rapid quality-control wine analyses. *Am J Enol Vitic.* 2003;**54**(2):131-134.

203. Regmi U, Palma M, Barroso CG. Direct determination of organic acids in wine and wine-derived products by Fourier transform infrared (FT-IR) spectroscopy and chemometric techniques. *Anal Chim Acta.* 2011;**732**:137-144. doi:10.1016/j.aca.2011.11.009.

204. Nieuwoudt HH, Pretorius IS, Bauer FF, Nel DG, Prior BA. Rapid screening of the fermentation profiles of wine yeasts by Fourier transform infrared spectroscopy. *J Microbiol Methods.* 2006;**67**(2):248-256. doi:10.1016/j.mimet.2006.03.019.

205. Rasines-Perea Z, Prieto-Perea N, Romera-Fernández M, Berrueta LA, Gallo B. Fast determination of anthocyanins in red grape musts by Fourier transform mid-infrared spectroscopy and partial least squares regression. *Eur Food Res Technol.* 2015;**240**(5):897-908. doi:10.1007/s00217-014-2394-6.

206. Jensen JS, Egebo M, Meyer AS. Identification of spectral regions for the quantification of red wine tannins with fourier transform mid-infrared spectroscopy. *J Agric Food Chem.* 2008;**56**(10):3493-3499. doi:10.1021/jf703573f.

207. Dambergs R, Gishen M, Cozzolino D. A review of the state of the art, limitations, and perspectives of infrared spectroscopy for the analysis of wine grapes, must, and grapevine tissue. *Appl Spectrosc Rev.* 2015;**50**:261-278. doi:10.1080/05704928.2014.966380.

208. Cocciardi RA, Ismail AA, Sedman J. 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.* 2005. doi:10.1021/jf048663d.

209. Llario R, Iñón FA, Garrigues S, De La Guardia M. Determination of quality parameters of beers by the use of attenuated total reflectance-Fourier transform infrared spectroscopy. In: *Talanta.* Vol 69. Elsevier; 2006:469–480. doi:10.1016/j.talanta.2005.10.016.

210. Shah N, Cynkar W, Smith P, Cozzolino D. Use of attenuated total reflectance midinfrared for rapid and real-time analysis of compositional parameters in commercial white grape juice. *J Agric Food Chem.* 2010;**58**:3279–3283. doi:10.1021/jf100420z.

211. Versari A, Parpinello GP, Scazzina F, Rio D Del. Prediction of total antioxidant capacity of red wine by Fourier transform infrared spectroscopy. *Food Control.* 2010;**21**(5):786–789. doi:10.1016/j.foodcont.2009.11.001.

212. Di Egidio V, Sinelli N, Giovanelli G, Moles A, Casiraghi E. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *Eur Food Res Technol.* 2010;**230**(6):947–955. doi:10.1007/s00217-010-1227-5.

213. Buratti S, Ballabio D, Giovanelli G, *et al.* Monitoring of alcoholic fermentation using near infrared and mid infrared spectroscopies combined with electronic nose and electronic tongue. *Anal Chim Acta.* 2011;**697**:67–74. doi:10.1016/j.aca.2011.04.020.

214. Preserova J, Ranc V, Milde D, Kubistova V, Stavek J. Study of phenolic profile and antioxidant activity in selected Moravian wines during winemaking process by FT-IR spectroscopy. *J Food Sci Technol.* 2015;**52**(10):6405–6414. doi:10.1007/s13197-014-1644-8.

215. Fayolle P, Picque D, Perret B, Latrille E, Corrieu G. Determination of Major Compounds of Alcoholic Fermentation by Middle-Infrared Spectroscopy: Study of Temperature Effects and Calibration Methods. *Appl Spectrosc.* 1996;**50**(10):1325–1330. doi:10.1366/0003702963904872.

216. Teixeira dos Santos CA, Páscoa RNMJ, Porto PALS, Cerdeira AL, Lopes JA. Application of Fourier-transform infrared spectroscopy for the determination of chloride and sulfate in wines. *LWT - Food Sci Technol.* 2016;**67**:181–186. doi:10.1016/j.lwt.2015.11.050.

217. Fragoso S, Aceña L, Guasch J, Mestres M, Busto O. Quantification of phenolic compounds during red winemaking using FT-MIR spectroscopy and PLS-regression. *J Agric Food Chem.* 2011;**59**(20):10795–10802. doi:10.1021/jf201973e.

218. Cozzolino D. State-of-the-art advantages and drawbacks on the application of vibrational spectroscopy to monitor alcoholic fermentation (beer and wine). *Appl Spectrosc Rev.* 2016;**51**(4):282–297. doi:10.1080/05704928.2015.1132721.

219. Santos PM, Pereira-Filho ER, Rodriguez-Saona LE. Application of hand-held and portable infrared spectrometers in bovine milk analysis. *J Agric Food Chem.* 2013;**61**(6):1205–1211. doi:10.1021/jf303814g.

220. Birkel E, Rodriguez-Saona L. Application of a portable handheld infrared spectrometer for quantitation of trans fat in edible oils. *JAOCS, J Am Oil Chem Soc.* 2011;**88**(10):1477–1483. doi:10.1007/s11746-011-1814-z.

221. Allendorf M, Subramanian A, Rodriguez-Saona L. Application of a handheld portable mid-infrared sensor for monitoring oil oxidative stability. *JAOCS, J Am Oil Chem Soc.* 2012;**89**(1):79–88. doi:10.1007/s11746-011-1894-9.

222. Sorak D, Herberholz L, Iwascek S, Altinpinar S, Pfeifer F, Siesler HW. New developments and applications of handheld raman, mid-infrared, and near-infrared spectrometers. *Appl Spectrosc Rev.* 2012;**47**(2):83-115. doi:10.1080/05704928.2011.625748.

223. Lachenmeier DW, Godelmann R, Steiner M, Ansay B, Weigel J, Krieg G. Rapid and mobile determination of alcoholic strength in wine, beer and spirits using a flow-through infrared sensor. *Chem Cent J.* 2010;**4**(1). doi:10.1186/1752-153X-4-5.

224. Dabros M, Genasci G, Blanchard L, *et al.* On-line monitoring and control of fed-batch fermentations in winemaking. *Chimia (Aarau).* 2016;**70**(12):900-901. doi:10.2533/chimia.2016.900.

225. Sablayrolles JM. Control of alcoholic fermentation in winemaking: Current situation and prospect. *Food Res Int.* 2009;**42**(4):418-424. doi:10.1016/j.foodres.2008.12.016.

226. Hernández G, León R, Urtubia A. Detection of abnormal processes of wine fermentation by support vector machines. *Cluster Comput.* 2016;**19**(3):1219-1225. doi:10.1007/s10586-016-0594-5.

227. Urtubia A, Pérez-Correa JR, Pizarro F, Agosin E. Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. *Food Control.* 2008;**19**:382-388. doi:10.1016/j.foodcont.2007.04.017.

228. Urtubia A, Hernández G, Roger JM. Detection of abnormal fermentations in wine process by multivariate statistics and pattern recognition techniques. *J Biotechnol.* 2011;**159**:336-341. doi:10.1016/j.jbiotec.2011.09.031.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chapter 2

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Hypothesis and objectives

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Attenuated Total Reflectance Mid-Infrared (ATR-MIR) spectroscopy has made it possible to analyse, in a fast and effective way, the composition of numerous food and beverage matrices. In the wine industry, it has been shown that ATR-MIR can be used to monitor several chemical parameters during alcoholic fermentation. However, its use as a Process Analytical Technology (PAT) tool has not been yet fully investigated in the literature.

In the wine industry, the advantages of implementing fast analytical tools along the key stages of wine production (maturation, fermentation and aging) are well known, but it would be important to test whether these technologies are able to detect different types of deviations that may arise during the fermentation process. In this thesis, we focused on the deviations associated with the fermentation process that can substantially affect the quality of the final wine (sluggish fermentations and bacterial spoilage). Because the ATR-MIR as a PAT tool to control wine fermentation is in its early stages of development, all of these hypotheses were tested in laboratory-scale fermentations.

Main objective

To evaluate the suitability of a portable ATR-MIR spectrometer as a PAT tool to monitor the alcoholic fermentation process of white grape must on a laboratory-scale. To this aim, different chemometric tools were evaluated. This main objective can be subdivided in the following specific objectives.

Specific objectives

- To evaluate the capacity of a portable ATR-MIR instrument to predict quality control parameters during all the stages of alcoholic fermentation using multivariate regression methods.

- To extract relevant information of the alcoholic fermentation process from ATR-MIR spectra, applying advanced chemometric techniques, including:

    - ANOVA-Simultaneous Component Analysis:
    - Multivariate Curve Resolution Alternating Least Squares.

- To detect deviations of normal operating conditions (NOC) fermentations using ATR-MIR spectra and multivariate classification methods. The induced deviations were:

  - Nitrogen deficiency to induce a sluggish fermentation.
  - Inoculation of bacteria to induce a bacterial spoilage:
    - Inoculation of lactic acid bacteria to induce malolactic fermentation.
    - Inoculation of acetic acid bacteria to induce and increase of volatile acidity.

- To develop control charts that could be used to detect deviations based on modelled NOC fermentations using techniques of multivariate data analysis.

Chapter 3

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Article 1

"Early detection of undesirable deviations in must fermentation using a portable FTIR-ATR instrument and multivariate analysis"

Published in Journal of Chemometrics, Volume 33, Issue 8, August 2019, e316

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

The use of portable IR spectrometers has opened the possibility of applying IR spectroscopy together with chemometrics to monitor manufacturing processes on-site, that is, directly into the field. These instruments, however, require a prior validation to be accepted as suitable tools. The next chapter tries to address the main objective of this thesis (first and second subobjectives), focusing on the application of a portable ATR-MIR instrument to monitor alcoholic fermentation.

First, we optimised the method of spectra acquisition in terms of sampling (sample pretreatment) and instrumental analysis (number of scans and resolution of the IR spectrometer). Alcoholic fermentation was monitored by analysing the fermenting samples at different time points during the course of fermentation.

Despite process data can be regarded as a 3D data matrix built from layers of batches, a simpler way to arrange the data is assessed based on a 2D array unfolding. Unfolding means that all original batch data blocks are concatenated vertically or horizontally and, in this way, PCA models can be built in different ways. First, global PCA models (horizontal concatenation) were built by considering each batch as an independent representation of the process. Alternatively, *k*-PCA models (where *k* refers to a specific time point in which spectra are acquired) do not consider the whole process, but allow evaluating the process status at a specific point. Spectral residuals and Hotelling T^2 were calculated, combined and simultaneously analysed in the so-called "influence plots", providing the most-used classical way to detect abnormal samples.

PLS regression was the chemometric tool used for quantitative determination of key fermentation monitoring parameters. Spectra were evaluated using a linear combination of the data with each individual chemical component, which is the basis for process control and monitoring.

One common feature of fed-batch processes is the unavoidable variation among batches due to natural biological variability and to differences in composition of raw materials, which causes the well-known "batch-effect" (also found as inter-batch effect). These effect is produced when the basic shape of the time trajectories from batch to batch is similar, but their time duration (between runs) changes. A novel approach called "biological process time", taking into account this variation within batches, is evaluated on fermentation samples.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Early detection of undesirable deviations in must fermentation using a portable FTIR-ATR device and multivariate analysis

Julieta Cavaglia[1]*| Barbara Giussani[2]| Montserrat Mestres[1]| Miquel Puxeu[3]|Olga Busto[1]| Joan Ferré[4]| Ricard Boqué[4]

[1]Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, iSens Group, Campus Sescelades, 43007 Tarragona, Spain

[2]Università degli Studi dell'Insubria, Dipartimento di Scienza e Alta Tecnologia, Via Valleggio 9, 22100 Como, Italy

[3]Wine Technology Park (VITEC), Falset, Tarragona, Spain

[4]Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, Chemometrics, Qualimetrics and Nanosensors Group, Campus Sescelades, 43007 Tarragona, Spain

*Corresponding author: julieta.cavaglia@urv.cat

Summary

A portable FTIR-ATR spectrometer was used to monitor small-scale must fermentations (microvinifications) with the aims to describe the process and to early detect problematic fermentations. Twenty fermentations at normal operation conditions (NOC) and 3 fermentations that were intentionally deviated from NOC (yeast assimilable nitrogen deficiency - YAN) were monitored. FTIR-ATR spectra were registered after a minimum sample pretreatment during the fermentation process. In addition, density, sugars (glucose and fructose) and acetic acid contents were determined by traditional methods.

Different multivariate analysis strategies (global and local models) were applied to the spectroscopic data to describe the evolution of the NOC fermentation and to early detect the abnormal fermentations. Global models based on principal component analysi (PCA) and partial least squares discriminant analysis (PLS-DA) allowed to describe the fermentations evolution in time and to correctly classify NOC and YAN fermentations. Abnormal deviations were successfully detected by developing one model for each sampling time. YAN experiments could be identified 49 hours after the beginning of the fermentations by means of Hotelling T^2 and residual F statistics.

In conclusion, ATR-FTIR coupled to multivariate analysis showed great potential as afast and simple at-line analysis tool to monitor wine fermentation and to early detect fermentation problems.

Key words: ATR-FTIR, fermentation monitoring, multivariate analysis, wine.

Introduction

In the winemaking industry, the control of the whole production chain, from harvest to bottling, is essential to obtain high-quality wines. One of the crucial phases in wine production is certainly the must fermentation, which is the biological transformation of grape juice into wine. Whereas it comprises many biochemical reactions, the most important change is the conversion of sugars into ethanol and CO_2. Nevertheless, the secondary reactions that take place during must fermentation have a substantial impact on the quality, flavor and character of the final wine[1].

Must fermentation requires, therefore, a thorough monitoring: failing to achieve a successful process control at this stage may result in stuck or sluggish fermentations that could throw away a whole vintage or lead to low quality wines[2].

Several routine measurements such as density, temperature and pH, are usually carried out throughout the fermentation process in wine cellars. However, additional measurements (e.g. total and volatile acidity, sugars, SO_2, assimilable nitrogen) which are often costly, time-consuming and require specific equipment and personnel, are commonly performed to gain more information[3].

In 2004 the United States Food and Drug Administration introduced the concept of 'Process Analytical Technologies' (PAT), aiming at implementing a real-time monitoring system through the production chain. This would replace final product testing as quality is controlled during the production process, giving the possibility to 'readjust' a process before the product is made and thus minimizing rejects[4].

Over the last decades, infrared spectroscopy, in combination with multivariate analysis, has proven to be a powerful tool for food analyses and, specifically, for wine analyses. Partial Least Squares Regression (PLSR) has been the most used calibration algorithm to predict chemical or physical parameters in wine from spectroscopic data[5].

As reviewed by dos Santos *et al.*, it has been shown that Near Infrared Spectroscopy (NIR) and Mid Infrared Spectroscopy (MIR) are both suitable techniques to predict several quality control parameters in grape juice, must and wine at different production stages, including total sugars (mainly glucose and fructose), ethanol, glycerol, total phenolics, anthocyanins or acetic acid, among other compounds[6]. The potential of NIR and MIR to monitor and model alcoholic fermentations was also investigated, demonstrating the usefulness of these techniques to monitor the evolution of the fermentation process[7-10].

Regmi *et al.* used MIR in the transmission mode with PLSR to predict the concentration of several acids in wine. They obtained good calibration results for citric, malic, tartaric, acetic, succinic, and lactic acids[11]. Moreover, MIR spectroscopy with PLS regression was also used for the quantification of reducing sugars, titrable acidity, total soluble solids, pH, and some phenolic compounds (see the review by Dambergs *et al.* and references therein)[12].

Among the different vibrational spectroscopic modes, the attenuated total reflectance MIR (ATR-MIR) mode is particularly advantageous over traditional transmission MIR modes because it requires little or even no sample pretreatment and it is faster and simpler to use. Moreover, as the infrared beam only penetrates the samples a few microns, so typical spectra saturation due to the high-water absorption band does not occur[13].

ATR-MIR was successfully employed to determine the total soluble solids (°Brix), pH, total phenolics, ammonia, free amino nitrogen, and yeast assimilable nitrogen (YAN) in grape juice samples[14]. Kim *et al.* were able to predict alcohol, reducing sugars and titratable acidity in fermenting samples of Makgeolli rice wine using ATR-MIR, thus proving the suitability of this technique to monitor the fermentation process[15]. Wu *et al.* used ATR-MIR to successfully monitor the course of Chinese rice wine fermentation[16].

The researchers were capable to predict total sugar, ethanol, titratable acidity, and amino nitrogen by applying different calibration models. Previously, Cozzolino *et al.* had also investigated the suitability of ATR-MIR to predict the time course of fermentation in samples at different days of fermentation using PLS discriminant analysis (PLS-DA) models. They obtained promising results, with low standard errors of prediction[17].

Portable FTIR instruments are rapidly gaining popularity across the food industry sector. They are cheaper, simpler to use, and faster than traditional instruments and allow sample analysis to be performed directly on the field: for these reasons, they could be considered powerful tools to rapidly perform quality control test and process monitoring especially when coupled with multivariate analysis. Portable FTIR instruments have been used for multiple purposes in foodstuff analysis, including, eg, the prediction of fatty acid content in marine oil, quantification of acrylamide in potato chips, or quantification of trans-fat content in fat and oil samples[18-20]. To our knowledge, this is the first time that a portable ATR-FTIR device is used for the analysis of must and wine fermenting samples.

The aim of this research was to develop a strategy to monitor the must fermentation and to early detect deviation from the typical fermentation using a portable ATR-FTIR

instrument coupled with multivariate analysis. The first step of the study concerned the investigation of the suitability of the instrument to the scope. Twenty-three must fermentations were carried out, and data were recorded during the whole process after a minimum sample pretreatment. Different multivariate approaches were applied for modeling the typical fermentation process, thus describing the normal operation conditions (NOC), and to early predict deviation from the NOC, in particular for a fermentation run with deficiency of assimilable nitrogen. The choice of the chemometric strategy was driven by the idea to give to winemakers a quite easy to understand process control model, which coupled with a portable device resulted in a process control methodology cheap and easy to implement.

Material and methods

Samples

Concentrated white natural must was obtained from "Concentrats Pallejà" (Riudoms, Spain). This was diluted 1:4 with distilled water to give an initial sugar (glucose and fructose) concentration of about 200 $g \cdot L^{-1}$ (to emulate the concentration of sugars found in a must coming from optimal mature grapes) and supplemented with 0.3 $g \cdot L^{-1}$ of Actimaxbio* (Agrovin, Ciudad Real, Spain) to ensure a YAN source. Table 1 summarizes the chemical parameters of must once diluted and supplemented.

Table 1. Chemical Parameters of diluted must.

209 $g \cdot L^{-1}$ glucose + fructose
228 $g \cdot L^{-1}$ yeast assimilable nitrogen
pH = 3.94
Total Acidity = 7.0 g Tartaric acid$\cdot L^{-1}$
Density = 1.0865 $g \cdot mL^{-1}$
Malic acid = 2.12 $g \cdot L^{-1}$

The microvinifications were conducted in 500 mL Erlenmeyer flasks containing 350 mL of diluted must and under constant temperature of 18°C. Twenty microvinifications were carried out without manipulating them or varying any parameter (NOC). Moreover, three microvinifications were intentionally altered to promote nitrogen deficiency: they were run without the addition of the YAN source.

Yeast and nutrients

The alcoholic fermentations were carried out by *Saccharomyces cerevisiae* yeast, and the inoculation was done as follows: 3.15 g of active dry yeast "VitilevureDV10" (Danstar Ferment AG, Denmark) was rehydrated in 60 mL of milliQ water, and 2 mL of yeast solution was added to the 23 Erlenmeyer flasks containing 350 mL of must, to reach a final concentration of 0.3 $g \cdot L^{-1}$ in each flask.

ATR-FTIR spectroscopic analysis

Data acquisition was performed using a portable 4100 ExoScan FTIR instrument (Agilent, California, USA), equipped with an interchangeable spherical ATR sampling interface, consisting on a diamond crystal window.

A total of 17 sampling points (times) were analyzed before the end of fermentation. Samples were randomly collected twice a day (every 12 hours approximately), centrifuged at 10 000 rpm for 10 minutes so that the supernatant could be collected using a micropipette. A drop of the supernatant was placed on top of the crystal using a Pasteur pipette, ensuring that the surface was completely covered with the sample, and the spectrum was recorded immediately afterwards. All spectra were recorded in the region of 3999 to 649 cm, with 32 scans and 8 cm^{-1} resolution. An air background was collected after every triplicate, that is, one background per sample. After each measurement, the crystal was carefully cleaned using deionized water and cotton wipes. Spectra were examined using the Microlab PC software (Agilent, California, USA), and data were saved as.spc files.

Absorbance data were used for the chemometric calculation. The mean of the sample replicates was calculated, and different preprocessing (smoothing and normalization) methods were tested in order to remove unwanted variations not due to changes in chemical compounds during fermentation, such as baseline drifts and noise observed in the raw spectra.

The final data was a three-way array containing the spectroscopic signals of 23 samples (20 NOC and three YAN) with 899 wavelengths recorded for 17 times covering a total of 258 hours of fermentation.

Quality Control Parameters

Reference analyses were carried out every 24 hours to monitor the fermentation process. Density was measured using an Densito 30PX electronic densimeter (Mettler Toledo),

whereas sugars (glucose and fructose) and acetic acid were determined using a Y15 Analyser (Biosystems, Barcelona, Spain). All the analyses were performed right after sample collection.

Multivariate Analysis

The collected data consisted of a three-way structure containing spectra (J = 899), batches or samples (I = 23), and sampling times (K = 17). Depending on the information we want to obtain, this data matrix can be treated as a multiway structure, unfolded into a two-way structure or divided into several matrices, usually one for each sampling time. Unfolding can be performed in several ways, depending on the mode that is kept in common. If unfolding is performed sample-wise, the final matrix has dimensions (I×KJ), with each row containing the spectra of a given sample at the different time points. If the spectral mode is common, then the final unfolded matrix has dimensions (J×IK). In this last matrix, each row contains a spectrum of sample i at time point k. Finally, if unfolding is performed timewise, the final matrix has dimensions (K×JI), where each row contains the spectra of all samples at time point k. Once unfolded, the matrix structure can be processed also in different ways. Global approaches can be applied, which means that all the data collected throughout the process are used in a global model. Alternatively, local approaches refer to the use of data separately from each sampling time to build independent models[21]. Principal component analysis (PCA), partial least squares regression (PLSR), and PLS-DA were used to process the data. The strategies used in this work are described in the following section.

All the models were cross-validated with random subsets (10 splits and five iterations). In PLSR and in PLSDA, the root mean square error of cross-validation (RMSECV) error was used to estimate the optimum number of latent variables to be used in prediction.

All multivariate data analyses were performed using the PLS Toolbox v8.6.1 (Eigenvector Research Inc., Eaglerock, USA) with MATLAB R2015b (The MathWorks, Natick, USA).

Results and discussion

Spectroscopic Data

Firstly, the signal quality was investigated. Several combinations of spectral resolution and number of acquisition scans were tested. An increase in the resolution (8-4-2 cm^{-1} was tested) did not add any relevant information to the spectra: peaks were well described, and this was confirmed by the chemometric modeling, which did not change in performances

when using spectra recorded at higher resolution values. Regarding data acquisition, scan numbers from 32 to 512 were tested, but the final models did not change relevantly in their performances. For this reason, a more rapid solution (32 scans) was preferred as it allowed reaching satisfactory results.

The evolution of the ATR-FTIR spectra throughout the whole fermentation process is shown in Figure 1. Due to the high absorbance of the O-H bond of water in the mid-infrared region and the high amount of overlapping vibrational modes in similar molecules, single molecules peak assignment is quite difficult. The main changes in the spectra are found between 950-1500 cm^{-1} and 3000-3500 cm^{-1}. The bonds in the 950-1500 cm^{-1} region could be associated with sugars and organic acids. Peaks between 1500 and 1200 cm^{-1} correspond mainly to deformations of $-CH_2$, deformations of C–C–H and H–C–O. On the other hand, peaks between 1200 and 950 cm^{-1} could be related to stretching modes of C–C and C–O. The broad band between 3000 and 3500 cm^{-1} could be ascribed to water and ethanol O-H stretching vibrations These results are in agreement with the literature, both in ATR and transmission IR modes.[22]

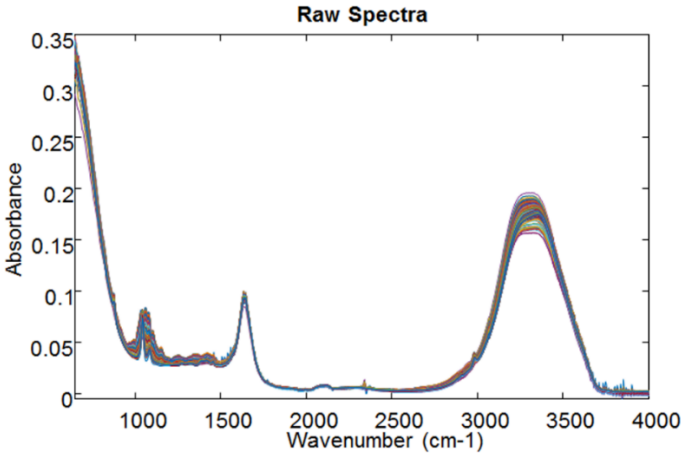


Figure 1. FTIR full spectra for all the fermenting samples (including all time points).

Data preprocessing

After calculation of the mean of the sample replicates, different preprocessing methods were tested to overcome baseline drifts and noise observed in the raw spectra. The following combination of preprocessing methods gave the best results:

- Smoothing (Savitzky-Golay) filter: window size 11pts, polynomial order 2
- Standard Normal Variate (SNV) normalization
- Mean Centering

Because the objective of the work was to detect deviations from the NOC, the average trajectory of each variable was subtracted in the batches. In this way, models focused the attention on the variability around these trajectories.

Fermentation control parameters

Density, sugars (glucose and fructose) and acetic acid values during fermentation are depicted in Figure 2, in which NOC samples are described by circles and YAN samples are indicated with stars. Density vary between 1,09 $g \cdot mL^{-1}$ at the beginning of the fermentation and 0,99 at the end of the process, showing typical values for white wine fermentations. NOC samples reached sugar depletion sooner than nitrogen-deficient samples. This behaviour could be explained considering that a lack of nutrients causes a decrease in yeast's enzymatic activity, which results in sluggish fermentations[2].

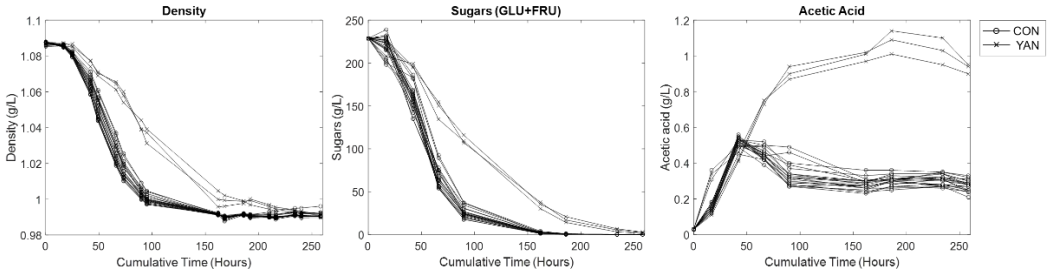


Figure 2. Evolution of chemical parameters: A: Density; B: Sugars (Glucose+Fructose); and C: Acetic acid.

A higher production of acetic acid could be observed in the nutrient deficient samples. Acetic acid is a by-product of yeast metabolism, which is generated from acetyl-coenzyme A derived from oxidative decarboxylation of pyruvate[23]. An increase of its values could be often observed in stuck fermentations, where conditions for yeast development are not optimal[24].

Global PCA model

First, we decided to explore the whole data set following a global approach. Data collected from NOC experiments were arranged in a two-way unfolded matrix with samples × times in the rows and spectra (wavenumbers) in the columns, with the aim to study the sample evolution throughout the fermentation process.

The final matrix had size 391 × 899. The score plot for the first two PCs (90.16% of the total variance) is reported in Figure 3. A trend in the samples position clearly emerges from the graph: samples are located along the first PC, from positive to negative values, according to the sampling time. All the NOC experiments and the YAN experiments showed a similar trend. While the PC1 accounted for the spectra variation in time, the second PC seemed to account for an experimental variability that could be possibly related to small differences between the evolving of samples during the fermentation process. Focusing the attention on PC1, the scores showed a tendency very similar to the one described for density and sugar values, confirming that this component mainly explains the fermentation evolution in time. Moreover, it is possible to distinguish the NOC and YAN fermentations that show a similar but not identical behavior. This model was able to detect the main changes in the samples at the different sampling times. This first promising result motivated us to further investigate the possibility to use the portable ATR-FTIR instrument to monitor the wine fermentation process.

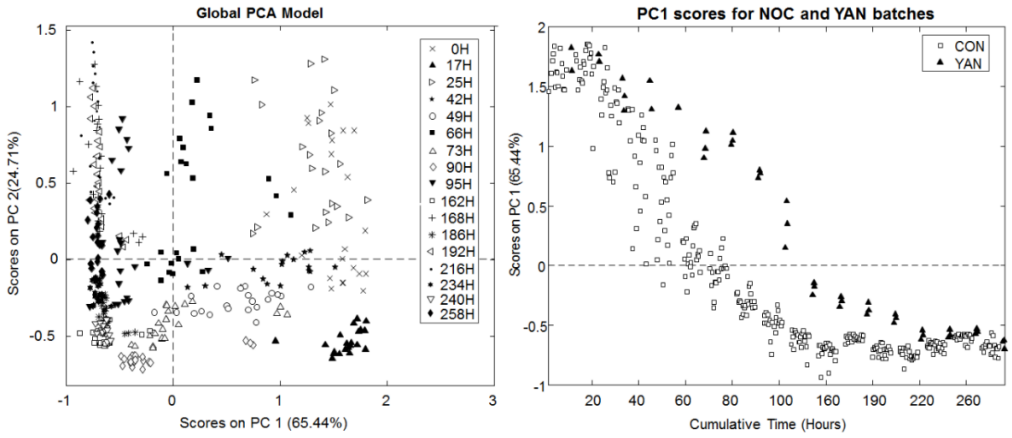


Figure 3. Scores plot for the global PCA model (left), samples are marked according to their sampling time. PC1 scores for NOC and YAN batches (right).

A partial least squares (PLS) regression model was then built on the same unfolded data matrix to predict the total sugars (glucose and fructose) concentration values from the recorded ATR-FTIR spectra along the fermentation. The values obtained with the reference analytical method were used as the Y data. The aim of this model was to prove the suitability of the portable ATR-FTIR spectrometer to monitor the wine fermentation through the prediction of one of the most important parameters, that is, the change in the total sugar content along fermentation. The statistical parameters of the regression model (two factors accounting the 98.68% of the Y variability) were RMSEC = 10.6 g·L^{-1}, RMESCV = 10.9 g·L^{-1}, R^2 = 0.987, and bias = −0.02 g·L^{-1}.

Figure 4 shows the measured vs PLSR predicted total sugar. There is a good agreement between measured and predicted values, confirming that coupling ATR-FTIR portable spectroscopy and multivariate analysis allowed to successfully monitor one of the major changes in fermenting wine samples and possibly the whole fermentation process.

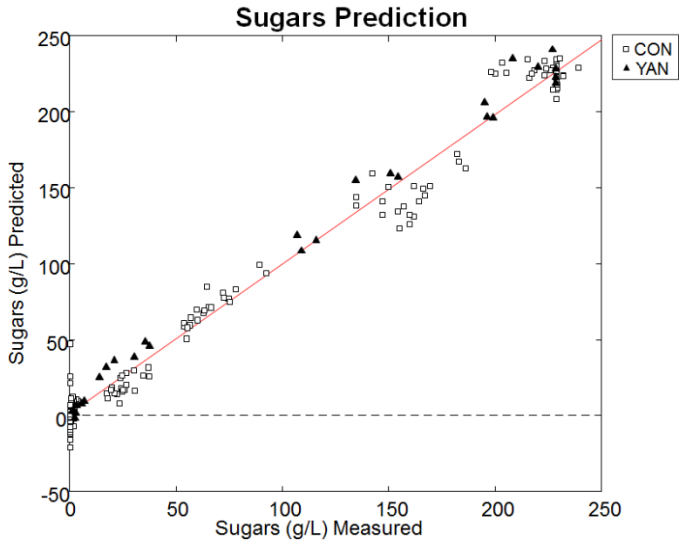


Figure 4. Measured vs Predicted concentrations of sugars (glucose+fructose).

Global PLS-DA model

The global data analysis strategy was then employed with the aim of evaluate the possibility to distinguish NOC fermentation from YAN fermentation using the spectra collected with the portable device during the whole fermentation process. In this case, the original three-way data matrix was unfolded in a time-wise manner so that sample direction was maintained. The unfolded matrix size was 23x15283 (23 samples x (899 variables x 17 time points)). A PLS-DA strategy was chosen due to the small number of samples and a PLS-DA model was built in order to classify fermentation experiments in NOC and YAN classes (in the Y vector, zeros were attributed to the NOC class samples, and ones were attributed to YAN class samples).

Figure 5 depicts the classification between normal and nitrogen deficient fermentations. As emerged from the graph, the two classes are well separated and no overlapping between them could be observed. The threshold used to discriminate between the classes was calculated as the value that best splits the classes with the least probability of both false

positives and false negatives (assuming that the predicted values for each class are approximately normally distributed). The algorithm is implemented in the PLS-Toolbox.

Even if the number of YAN fermentation experiments is quite small with respect to the NOC fermentation, these results are really promising, showing the possibility to distinguish the different types of fermentation when spectra collected along all the fermentation process are available.

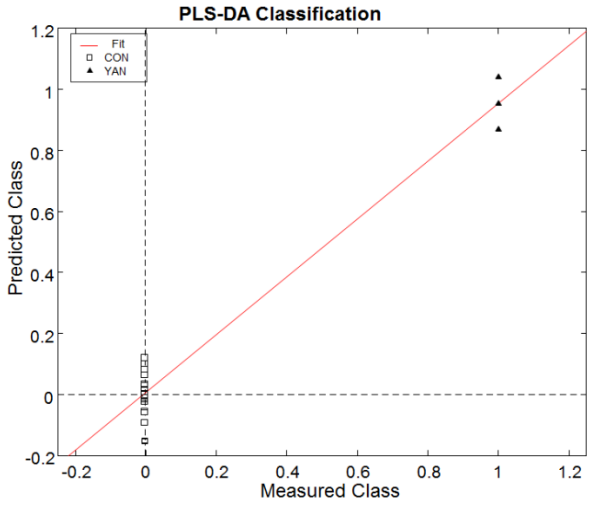


Figure 5. PLS-DA model for control (CON) and nutrient deficient samples (YAN). Zero was assigned to CON samples whereas ones was used for YAN samples.

k-PCA (Local Models)

A local strategy to early predict deviations from NOC was then developed. Local k-PCA models were built using the two-way matrices (samples × wavelengths [23 × 899]) obtained separately for each sampling time collected (a total of 17 data matrices, one for each time). A very satisfactory result was obtained, as the model built with spectra recorded after 49 hours (time point 4) was able to distinguish between NOC and YAN fermentations processes. Figure 6 shows the influence plot for PC1. The same result was obtained with the PLS-DA modeling strategy as expected. Several PLS-DA models were built, one for each sampling time. The PLS-DA model built after 49 hours (time point 4) gave the 100% of correct classification with no overlap between the classes (0 was attributed to NOC class, 1 was attributed to YAN class).

Using a moving window approach (see the article by Camacho *et al.* and references therein[21]) to try to perform an earlier prediction of the deviation from NOC did not provide

better results. A possible explanation to this behavior could be the quite small number of sampling points analyzed at the beginning of the fermentation, which is clearly the moment of the whole process in which the main changes (especially in abnormal fermentations) occurred. For this reason, any other evolving modeling approach was not considered in this first step of the research project.

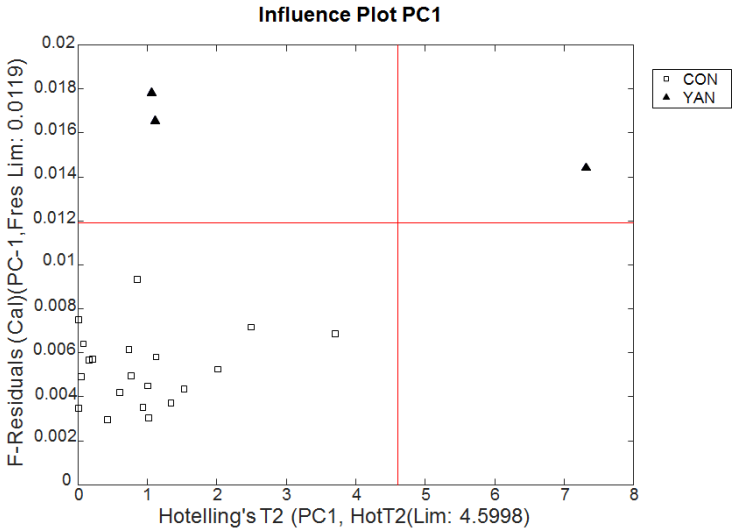


Figure 6. Influence plot for the k-PCA model at time point 4.

Biological process time

To monitor the evolution of the abnormal YAN fermentation, the approach developed by Jørgensen *et al.* was applied[25]. The reasoning behind the method is that each fermentation, starting similar initial conditions, can evolve slower or faster, and this different behavior can be detected. The idea is that spectra of the NOC samples can be modelled against the evolving time, but if this relationship is different for the abnormal batches then it means the fermentation has a different speed or has followed another direction. The method operates as follows:

1) The original data structure is unfolded keeping as common the spectral mode. Then, the relative times of all fermentations of all NOC samples are calculated, as the real time at time point *k* divided by the total time of the fermentation:

$$\text{Rel time}_{\text{ferm i}} = \frac{\text{Actual time}_{\text{ferm i}}}{\text{Total time}_{\text{ferm i}}}$$

The final time of a fermentation is assumed to have a relative time of 1, and the rest of relative times take values within 0 and 1. The relative time is also the % of evolution of the fermentation (relative time 0.6 means the fermentation is at 60%). Finally, a PLS regression model is built between the spectra of all NOC samples against the relative times. At this point, it is important to decide what the total time of a fermentation is. We decided to use the time where the sugar value was around the detection limit of the instrument, what coincided with the usual glucose value of a wine at the end of the fermentation process.

2) The spectra of all NOC samples at all fermentation times are regressed onto the previous PLS model to estimate what is called the "biological" process time. This is done because the assumption is that the difference between relative and biological time is due to the fermentation process.
3) A second PLS model is built between the NOC spectra and the "biological" time, that is, the predicted time of the first PLS model.
4) From this second PLS model, the resulting scores are used to build control charts for future batches. In these control charts (one for each PLS factor) confidence limits are calculated from the NOC training set (±2 and ±3 standard deviation curves) and represented *vs* "biological" time (see Fig 7).
5) Finally, to monitor future batches, their spectra are used in the second PLS model to predict the scores and the biological process time. Both predicted biological process time and scores are used in the control chart evaluations (see Fig 7). This allows on-line monitoring of batch evolution.

The approach was applied to monitor both normal control samples (NOC) and the YAN abnormal samples. Results areshown in Figure 7. It can be seen that YAN samples evolve in a substantially slower way, but the relationship betweenthe spectra and time works. The prediction of the biological time for the YAN sample confirms that, when the NOC samples are 100% fermented, YAN samples are about 60% fermented.

Conclusions

Monitoring the fermentation process is a crucial step in order to obtain high-quality wines and avoid materials and money waste. Several analytical techniques measuring a variety of analytes and properties fit for the purpose and give good performances, but often they need intensive sample preparation, or highly specialized instruments and operators, besides costly and time-consuming analyses. This work was focused on the use of a portable, easy-

to-use ATR-MIR device, coupled with multivariate analysis, as a rapid and economical strategy to monitor fermentation processes and to detect deviation from NOC.

The results obtained were very satisfactory. The prediction of the sugar content in fermenting samples from the beginning to the end of fermentation was performed, demonstrating the possibility to use this portable device to rapidly monitor fermentations running under normal operation condition. Moreover, slower fermentations (YAN) could be detected at an early stage of fermentation (when NOC are well described), giving the possibility to the winemaker to eventually correct the process and to obtain a good quality product.

Future work will be done increasing the number of samples both in NOC and in abnormal operation conditions, especially at the beginning of the fermentation, as it emerged from the models that the first 50 hours of fermentation are possibly the crucial ones to detect deviations from NOC conditions. We will take advantage of other strategies (eg, time evolving and moving average) to develop multivariate models. Moreover, a chemometric strategy will be developed to compare fermentations running in different times, for different wine types and including other problems that may occur during the fermentation process.

Acknowledgments

We thank the Spanish Ministry of Economy and Competitiveness (project AGL2015-70106-R) for financial support and AGAUR (Generalitat de Catalunya) for awarding the FI grant (2018 FI_B 00844) to Julieta Cavaglia.

References

1. Ciani M, Comitini F, Mannazzu I. In: Jørgensen SE, Fath BD, eds.Encyclopedia of Ecology. Oxford, UK: Elsevier B.V.; 2008:1548-1557.https://doi.org/10.1093/aob/mcp308.
2. Bisson LF. Stuck and sluggish fermentations.Am J Enol Vitic. 1999;50:107-119.
3. Ribéreau-Gayon P, Dubourdieu D, Donèche B, Lonvaud A.Handbook of Enology Volume 1 The Microbiology of Wine and Vinifications. 2nded. Chichester, UK: John Wiley & Sons Ltd; 2006.
4. Van Den Berg F, Lyndgaard CB, Sørensen KM, Engelsen SB. Process analytical technology in the food industry *.Trends Food SciTechnol. 2013;31(1):27-35.

5. Bauer R, Bauer FF, Kossmann J, Koch KR, Esbensen KH. FTIR spectroscopy for grape and wine analysis.Anal Chem. 2008;80(5):1371-1379.
6. dos Santos CAT, Páscoa RNMJ, Lopes JA. A review on the application of vibrational spectroscopy in the wine industry: from soil to bottle.Trends Anal Chem. 2017;88:100-118.
7. Di Egidio V, Sinelli N, Giovanelli G, Moles A, Casiraghi E. NIR and MIR spectroscopy as rapid methods to monitor red wine fermenta-tion.Eur Food Res Technol. 2010;230(6):947-955.
8. Urtubia A, Pérez-Correa JR, Pizarro F, Agosin E. Exploring the applicability of MIR spectroscopy to detect early indications of wine fer-mentation problems.Food Control. 2008;19(4):382-388.
9. Buratti S, Ballabio D, Giovanelli G, *et al.* Monitoring of alcoholic fermentation using near infrared and mid infrared spectroscopies com-bined with electronic nose and electronic tongue.Anal Chim Acta. 2011;697(1-2):67-74.
10. Emparán M, Simpson R, Almonacid S, Teixeira A, Urtubia A. Early recognition of problematic wine fermentations through multivariatedata analyses. Food Control. 2012;27(1):248-253.
11. Regmi U, Palma M, Barroso CG. Direct determination of organic acids in wine and wine-derived products by Fourier transform infrared(FT-IR) spectroscopy and chemometric techniques.Anal Chim Acta. 2011;732:137-144.
12. Dambergs R, Gishen M, Cozzolino D. A review of the state of the art, limitations, and perspectives of infrared spectroscopy for the anal-ysis of wine grapes, must, and grapevine tissue.Appl Spectrosc Rev. 2015;50(3):261-278.
13. Craig AP, Franca AS, Irudayaraj J.High Throughput Screening for Food Safety Assessment. Amsterdam, The Netherlands: Elsevier Ltd;2015:165-194.https://doi.org/10.1016/B978-0-85709-801-6.00007-1.
14. Shah N, Cynkar W, Smith P, Cozzolino D. Use of attenuated total reflectance midinfrared for rapid and real-time analysis of composi-tional parameters in commercial white grape juice.J Agric Food Chem. 2010;58(6):3279-3283.
15. Kim DY, Cho BK, Lee SH, Kwon K, Park ES, Lee WH. Application of Fourier transform-mid infrared reflectance spectroscopy for mon-itoring Korean traditional rice wine'Makgeolli'fermentation.Sens Actuators B. 2016;230:753-760.1

16. Wu Z, Xu E, Long J, *et al.* Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy.Food Control. 2015;50:405-412.
17. Cozzolino D, Curtin C. The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments.FoodControl. 2012;26(2):241-246.
18. Ayvaz H, Rodriguez-Saona LE. Application of handheld and portable spectrometers for screening acrylamide content in commercialpotato chips.Food Chem. 2015;174:154-162.
19. Karunathilaka SR, Mossoba MM, Chung JK, Haile EA, Srigley CT. Rapid prediction of fatty acid content in marine oil omega-3 dietarysupplements using a portable Fourier transform infrared (FTIR) device & partial least-squares regression (PLSR) analysis.J Agric FoodChem. 2017;65(1):224-233.
20. Mossoba MM, Kramer JKG, Azizian H, *et al.* Application of a novel, heated, nine-reflection ATR crystal and a portable FTIR spectrometerto the rapid determination of total trans fat.J Am Oil Chem Soc. 2012;89(3):419-429.
21. Camacho J, Picó J, Ferrer A. The best approaches in the on-line monitoring of batch processes based on PCA: does the modelling struc-ture matter?Anal Chim Acta. 2009;642(1-2):59-68.
22. Cozzolino D, Cynkar W, Shah N, Smith P. Feasibility study on the use of attenuated total reflectance mid-infrared for analysis of compo-sitional parameters in wine.Food Res Int. 2011;44(1):181-186
23. Aranda A, Matallana E, del Olmo M.Molecular Wine Microbiology. Amsterdam, The Netherlands: Elsevier Inc.; 2011:1-31. 10.1016/B978-0-12-375021-1.10001-3.
24. Alexandre H, Charpentier C. Biochemical aspects of stuck and sluggish fermentation in grape must.J Ind Microbiol Biotechnol.1998;20(1):20-27.
25. Jørgensen P, Pedersen JG, Jensen EP, Esbensen KH. On-line batch fermentation process monitoring (NIR)—introducing‘biological pro-cess time’.J Chemometr. 2004;18(2):81-91

Chapter 4

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

In this chapter, subobjectives three and four were addressed. The use of ATR-MIR together with chemometrics to detect bacterial spoilage during alcoholic fermentation was investigated. To this aim, a deeper understanding of the ATR-MIR spectra of wine was needed.

Finding the relevant MIR spectral bands affected by lactic acid bacteria spoilage was a difficult task. In this regard, ASCA was a suitable method to unravel small differences in the spectra coming from this spectral source of variability. The preliminary study entitled "*ASCA: a suitable exploratory tool to unravel small differences in spectroscopic data during wine alcoholic fermentation*", gave some insights on how an exploratory tool can serve as a basis for a variable selection strategy. In addition, testing different preprocessing strategies and determining the best combination of variables for the detection of bacterial spoilage, allowed us to monitor the time course of malolactic fermentation at different points of the alcoholic fermentation. The detection of contaminated fermentations was evaluated using the Hotelling T^2 and Q-residual statistics.

As in the previous chapter, the different arrangement of the spectroscopic data matrix allowed us to obtain different models for the detection of abnormal fermentations. In this case, PLS-DA was used to discriminate between normal and deviated samples.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Article 2

"ASCA: a suitable exploratory tool to unravel small differences in spectroscopic data during wine alcoholic fermentation"

To be submitted to the multivariate ANOVA special issue of Journal of Chemometrics, August 2021

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

ASCA: a suitable exploratory tool to unravel small differences in spectroscopic data during wine alcoholic fermentation

Julieta Cavaglia[a] | Barbara Giussani[b] | Montserrat Mestres[a] | Jean-Michel Roger[c,d] | Ricard Boqué[e]*

[a.] Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

[b.] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio, 9, Como (22100), Italy

[c.] ITAP, Univ Montpellier, INRAE, Institut Agro, 34196 Montpellier, France

[d.] ChemHouse Research Group, 34196 Montpellier, France

[e.] Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

*Corresponding author: ricard.boque@urv.cat

INTRODUCTION

In recent years, vibrational spectroscopic techniques, including Raman, Near Infrared (NIR) and Mid-Infrared (MIR) spectroscopy, have been gaining popularity for the monitoring and control of food-related processes [1]. These methods are considered fast-analytical tools, as they offer the possibility of obtaining information from a sample almost immediately [2].

In the wine industry, a correct monitoring of alcoholic fermentation is crucial in order to ensure the production of high-quality wines, as an inefficient management of this process would result in the production of undesired molecules having a negative effect on the quality of the final wine, both organoleptically and chemically speaking [3]. Traditionally, alcoholic fermentation is controlled by the daily measurement of chemical and physical parameters that ensure that the process is undergoing under the desired conditions [4]. However, these measurements are time-consuming, require specific laboratory equipment and trained personnel, making this system impractical for some wineries, which must send samples to external laboratories during the production process. The implementation of Process Analytical Technologies (PAT), which support the idea of controlling the quality of a product during processing, would be highly beneficial for those wineries that cannot afford having an own analytical laboratory in the field to perform at-line analyses. Even for wineries having their own laboratory, PAT strategies may be helpful, as the frequency of

analysis would increase substantially and pre-treatment of the samples would be considerably simplified [5].

Despite the clear advantages of using vibrational spectroscopy techniques as part of the PAT strategy for process monitoring (e.g. fast, easy-to-use techniques and minimum sample pretreatment), they require the application of chemometric tools, such as Partial Least Squares Regression (PLSR) or Partial Least Squares- Discriminant Analysis (PLS-DA) in order to extract useful information from spectroscopic data [6].

The first step in any study involving multivariate data analysis (MVDA) is data selection and pre-processing, which is usually the most time-consuming step [7]. This includes the selection of the most representative samples, the removal of redundant variables and, depending on the analytical instrument used, the use of data correction methods such as peak alignment for chromatographic data or baseline correction for spectroscopic data [8]. The type of preprocessing will largely depend on the aim of the study. In MIR data from alcoholic fermentation, it is almost mandatory to select the most important variables related to the molecules that are of interest. These variables are usually present in the fingerprinting region, where most organic bonds absorb infrared radiation [9]. In addition, the perfect pre-treatment combination (i.e. 1st or 2nd derivatives, Savintzky-Golay smoothing, multiplicative scatter correction) will strongly depend on the information related to the experimental problem at hand. This process also requires a vast knowledge from our data.

During alcoholic fermentation, the transformation of sugars (glucose and fructose) into ethanol and CO_2 accounts for most of the variability in the MIR spectra [10]. Although this is valuable information, it hinders the detection of other minor sources of variability, such as the production of organic acids that might be detrimental for the wine. An anomalous increase of acetic acid in the wine must (>0.5 $g \cdot L^{-1}$) suggests that acetic acid bacteria are developing in the wine and corrective measures must be taken immediately. Similarly, the production of lactic acid is a bad indicator of quality in most white wines, as it has a negative impact on their organoleptical profile [11].

In the food industry, when using vibrational spectroscopy for process control, it is common that the phenomenon we are interested in is hidden by unimportant variation from the main process [12]. It is in this type of situations where ANOVA Simultaneous Component Analysis (ASCA) can quickly help in finding other sources of variability different from the predominant ones. Applying ASCA before data preprocessing can drastically reduce the

time usually required to find the perfect preprocessing strategy for the data, as it would be possible to detect, directly from the raw spectra, what variables account for most of the variation for the effect(s) of interest [13]. In this way, during alcoholic fermentation, we might be interested in minor changes in the wine matrix, such as the production of organic acids, instead of the main process, which is the transformation of sugars into ethanol and CO_2 [14].

Applying ASCA to NIR spectral data from a bread staling process, Amigo *et al.* studied the contribution of the three major design parameters: enzyme treatment, measurement zone and storage time manifest. They found that the main effect, accounting for 73% of the variance in the data, was the measurement zone, while the spectral variance due to the time of storage (days) and the treatment accounted for 6.8% and 5.4%, respectively. The three main effects were found to be significant [15]. In another study, Grassi *et al.* studied the effect of yeast strains, temperature and fermentation time points on the variability in fermentation metabolites during beer fermentation. They suggested the use of interval-ASCA (i-ASCA), splitting variables into intervals of equal size in which each interval is independently evaluated. They found that time had always a significant effect in all intervals. The temperature and yeast strain factors showed significant influence ($p<0.01$) in some of the i-ASCA intervals, unlike classical ASCA results, which did not report significance for these factors [16].

The aim of this paper is to illustrate how ASCA is able to quickly unravel, from raw MIR data, the variation associated to the production of acetic and lactic acids during the alcoholic fermentation of white wine. The use of ASCA before data pre-processing and variable selection allowed to quickly address the regions of interest in the spectra that are related to the changes under investigation during alcoholic fermentation.

MATERIAL AND METHODS

Concentrated white must was provided by Mostos Españoles S.A., (Ciudad Real, Spain) and it was stored at -20 ºC until its use. Its defrosting was done at 5 ºC and it was then diluted with MilliQ water to adjust the sugar concentration to 200 ± 10 $g \cdot L^{-1}$. The diluted must was supplemented with 0.30 $g \cdot L^{-1}$ of ENOVIT® (SPINDAL S.A.R.L. Gretz Armainvilliers, France) and 0.30 $g \cdot L^{-1}$ of Actimaxbio* (Agrovin, Ciudad Real, Spain) in order to ensure a sufficient final concentration of yeast assimilable nitrogen. The commercial dry *Saccharomyces cerevisiae* yeast strain used was "E491" (Vitilevure Albaflor, YSEO, Danstar

Ferment A.G., Denmark). Regarding to lactic acid bacteria, a commercial freeze-dried blend of *Oenococcus oeni* and *Lactobacillus plantarum* "Co-inoculant Bacteria 3.2" (Anchor Oenology, South Africa) was used. Rehydration of the microorganisms was done following the suppliers' instructions.

For each fermentation batch, 350 mL of diluted must were added into 500 mL Erlenmeyer flasks and they were inoculated with 0.105 g of active dry yeast rehydrated in 2 mL of MiliQ water for 30 minutes at 25 ºC, reaching a final concentration of $3 \cdot 10^6$ CFU·mL^{-1}. In eight of the batches a lactic acid bacteria contamination was induced. LAB co-inoculations were done taking into account the producer instructions (1 g = $1 \cdot 10^{11}$ CFU·mL^{-1}), reaching two final concentrations: $2.5 \cdot 10^6$ (LAB1) and $4 \cdot 10^6$ CFU·mL^{-1}. To simulate the contaminations of the acetic acid bacteria, 5 microvinifications were spoiled by adding a medium consisting on *Acetobacter pasteurianus*, an acetic acid bacteria autochthonous strain, from the Oenological Biotechnology Department at Universitat Rovira i Virgili (Spain) to a final concentration of $1 \cdot 10^6$ CFU·mL^{-1}.

All microvinifications were kept under a constant temperature of 18 ºC until the end of alcoholic and malolactic fermentations. Alcoholic fermentation was considered finished when density was under 0.995 g·L^{-1}, whereas malolactic fermentation ended when L-malic acid concentration was < LOD (0.06 g·L^{-1}).

The dimensions of the final data matrix were 692 rows (wavenumbers x time points) and 266 variables (wavenumbers ranging from 1839 cm^{-1} to 850cm^{-1}), which consisted on 15 Normal Operation Conditions (NOC) and 13 abnormal microvinifications (5 inducing acetic acid bacteria spoilage (AAB) and 8 inducing lactic acid bacteria spoilage (LAB)).

In this dataset, three factors were considered to affect the spectra: time (fermentation progress), contamination (whether it was a normal or a deviated fermentation) and experiment (two independent experiments were conducted). The experiment factor was considered because of the "batch-effect" (also known as inter-experiment effect). One of the experiments consisted on 37 time points, while the other consisted on 8 time points. At the time of analysis, AAB fermentations were considered finished when the production of acetic acid reached its maximum (approx. 1.6 g·L^{-1}). Density measurements were made with a portable densimeter (Densito2Go, Mettler Toledo, United States). Reducing sugars, acetic acid and L-malic acid were measured using a Y15 Analyser (Biosystems, Barcelona, Spain). Acetic acid and sugars were measured until their concentrations were under 0.05 g·L^{-1}.

Density was measured to a minimum value (0.99 $g \cdot L^{-1}$), at which time fermentation was considered completed. All the analyses were performed right after sample collection.

After data preprocessing, ASCA models (were built using Matlab R2015 (The MathWorks, Natick, USA) and PLS Toolbox v8.7 (Eigenvector Research Inc., Eaglerock, USA). Four 60-hour time intervals were used to describe the different stages of alcoholic fermentation. In a batch process, when applying ASCA to different time intervals this is known as interval-ASCA models. Validation was performed using permutation tests and assessing the statistical significance (p-value). 500 permutations were applied and results were considered to be statistically significant when p-value <0.05.

RESULTS AND DISCUSSION

Figure 1 shows the evolution of the chemical parameters measured. In both experiments, lactic and acetic acid bacteria were capable to deviate the fermentation process through the production of lactic acid and acetic acid, respectively. In addition, a slight difference in the sugar consumption rate can be seen for AAB fermentations in comparison to NOC samples, whereas no difference between NOC and LAB samples can be appreciated. This suggests acetic acid bacteria are more stressful to yeast metabolism than lactic acid bacteria. The reason for this might be that the starting culture used to induce the LAB contaminations was especially indicated to be used for co-inoculation with yeasts, and these microorganisms are specifically selected to grow at the same time as yeasts.

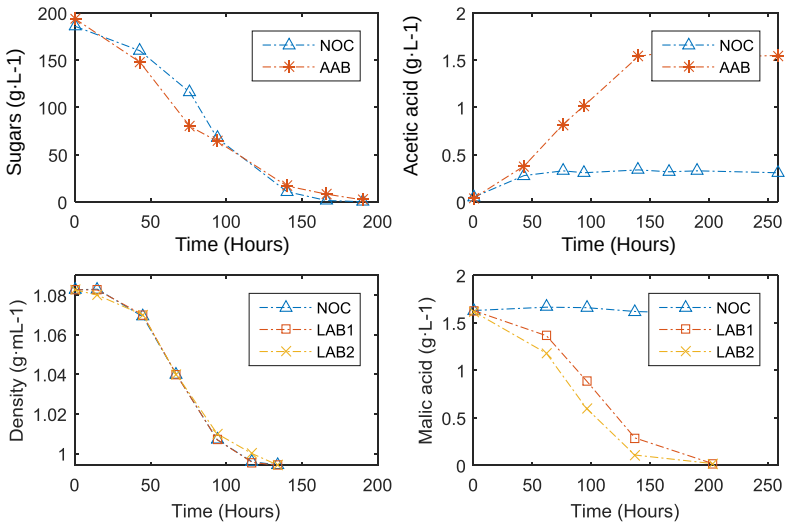


Figure 1. Evolution of chemical parameters during alcoholic fermentation.

Spectra preprocessing consisted of a Savitzky-Golay smoothing with a 15 points window and a 1st order polynomial, as well as Standard Normal Variate. This preprocessing reduced spectral noise and allowed a better chemical interpretation of ASCA loadings.

Table 1 shows the results for the ASCA models in all the fermentation intervals. The percentage of total variance accounting for each factor is indicated, as well as the *p*-value obtained after the permutation test.

Table 1. ASCA results showing the percentage of variance for each factor and the result of the permutation tests.

	Time interval 0-60h		Time interval 61-120h		Time interval 121-180h		Time interval 181h-End	
Factor	% Variance	p-value	% Variance	p-value	% Variance	p-value	% Variance	p-value
Time	89.1	0.001	92.9	0.001	20.4	0.001	20.2	0.001
Experiment	0.6	0.001	3.4	0.001	37.1	0.001	10.9	0.001
Contamination	0.1	0.538	0.1	0.014	11.1	0.001	9.8	0.001
Residuals	10.3		3.5		31.5		59.1	

At the beginning of the fermentation, both time and experiment are statistically significant (*p*-values <0.05) and account for 89.1% and 0.6% of the total variance. The contamination factor, however, is not significant. At 60 hours, the production of lactic acid and acetic acid is less than 0.8 $g \cdot L^{-1}$, which is close to the limit of detection of the instrument. In the second time interval (121-180 hours) the contamination factor has a statistical significance, but accounts for a very low variance in the spectra (0.1%) in comparison to time (92.9%) and experiment (3.4%). From the beginning of the fermentation to 120h, massive changes in the spectra occur due to the consumption of sugars by yeasts. In contrast, in the last interval the % variance for the factor "time" reduces drastically (20.2%), as sugar concentration after 180 hours is less than 10 $g \cdot L^{-1}$, whereas the % variance for the factor "contamination" does not vary significantly (9.8%). The residual matrix of the last interval represents 59.1% of the total variance. This matrix includes all the variations that the model cannot explain. The microvinifications conducted in this study were not prevented against the action of microorganisms, and the yeasts were not subtracted from the medium. Yeast lysis may explain this increase in the residuals, as well as the action from other microorganisms that could have grown after the end of alcoholic fermentation if residual sugars are present.

For illustration, the evolution of the % variance for the factor "time" is shown in Figure 2. It is interesting to note that the shape of the curve is very similar to that of a typical fermentation curve of sugar consumption.

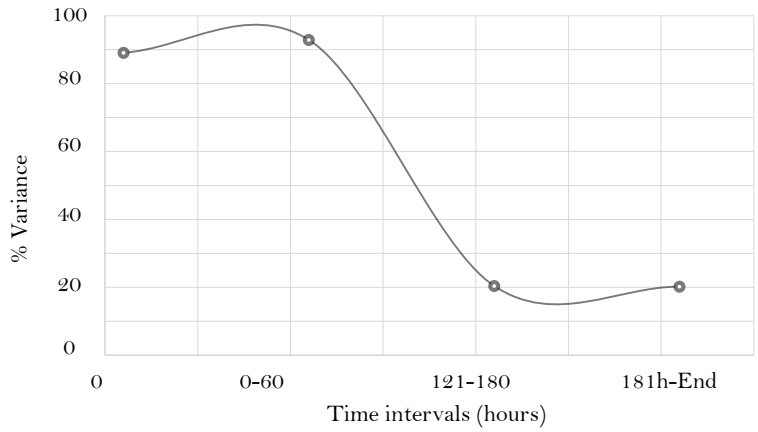


Figure 2. The evolution of % variance for the time factor along the different intervals of the fermentation.

One of the best features of ASCA is that a PCA can be built with the submatrices of interest, showing how the samples (scores) are distributed in the space according to those variables (loadings) that have more weight in the principal in of the PCA. Figure 3 shows the PCA of the contamination factor, for the time interval 121-180h, in which an 11.1% of the total variance of the data is attributed to the contamination factor. In this PCA, a clear grouping of the data can be observed.

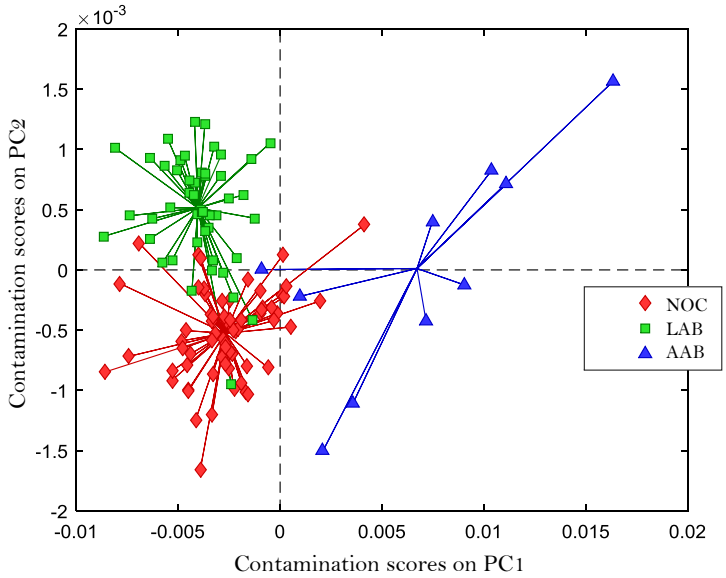


Figure 3. PCA plot for the contamination factor. NOC: fermentations in Normal Operation Conditions. LAB: Contaminated samples with Lactic Acid Bacteria. AAB: Contaminated samples with Acetic Acid Bacteria.

The loadings associated to the contamination factor shown in Figure 4. The arrows indicate those wavenumbers that can be clearly associated with the contamination factor. The wavenumbers showing clear peaks are 1720 cm^{-1}, 1268 cm^{-1}, 1196 cm^{-1} and 1134 cm^{-1}. These regions have already been associated with the absorption of organic acid bonds: the C=O bond stretching vibration of carboxylic acids shows an important peak at 1740 cm^{-1}. The O-H bending of the carboxylic acid and the C-O stretching absorb between 1200 and 900 cm^{-1}. Similarly, C-H bending vibrations absorb around 1400 and 1300 cm^{-1}.

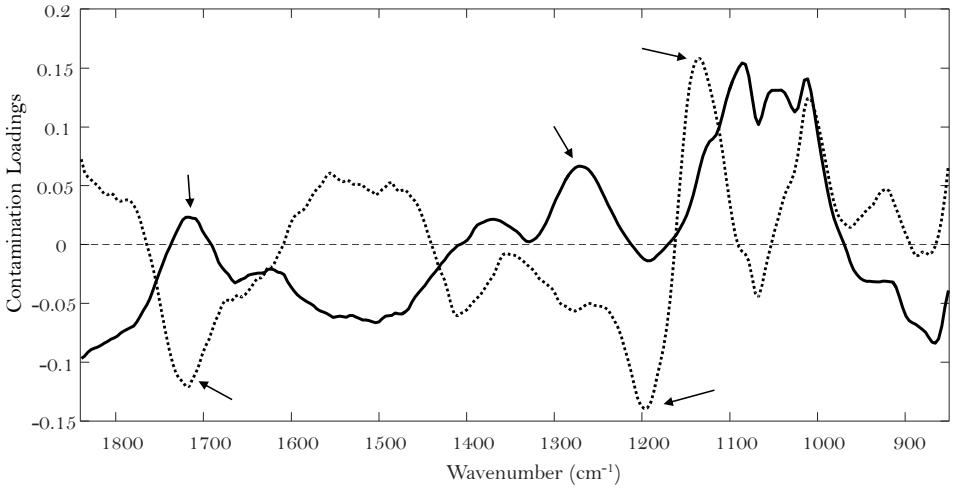


Figure 4. Loadings for PC1 (solid line) and PC2 (dotted line) of the PCA based on the contamination factor.

CONCLUSIONS

ASCA allows the quick determination of variable contributions to one or more experimental factors. In this preliminary study, we showed how the variance of the ATR-MIR spectra is affected along the alcoholic fermentation process, considering three of the most influencing factors in the process: time, contaminations and experiment. The application of ASCA before the development of any chemometric model allows optimising the selection of variables to be included in the analysis and determine how each of the evaluated factors in an experiment are affecting the measured variables. When analysing spectroscopic data, most of the times the factors of interest account for a very small percentage of the total variability. Here we have shown how ASCA is capable to detect a small variation due to lactic acid bacteria spoilage during wine alcoholic fermentation.

REFERENCES

[1] A.P. Craig, A.S. Franca, J. Irudayaraj, Vibrational spectroscopy for food quality and safety screening, in: High Throughput Screen. Food Saf. Assess. Biosens. Technol. Hyperspectral Imaging Pract. Appl., 1st Editio, Elsevier Ltd, 2015: pp. 165–194. https://doi.org/10.1016/B978-0-85709-801-6.00007-1.

[2] H. Huang, H. Yu, H. Xu, Y. Ying, Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review, J. Food Eng. 87 (2008) 303–313. https://doi.org/10.1016/j.jfoodeng.2007.12.022.

[3] L.F. Bisson, Stuck and sluggish fermentations, Am. J. Enol. Vitic. 50 (1999) 107–119.

[4] G. Specht, Yeast fermentation management for improved wine quality, in: Manag. Wine Qual., Elsevier, 2010: pp. 3–33. https://doi.org/10.1533/9781845699987.1.3.

[5] A. Urtubia, J. Ricardo Pérez-Correa, F. Pizarro, E. Agosin, Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems, Food Control. 19 (2008) 382–388. https://doi.org/10.1016/j.foodcont.2007.04.017.

[6] C.E. Miller, Chemometrics in Process Analytical Technology (PAT), 2010. https://doi.org/10.1002/9780470689592.ch12.

[7] L. Chuen Lee, C.-Y. Liong, A. Aziz Jemain, A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum, Chemom. Intell. Lab. Syst. 163 (2017) 64–75. https://doi.org/10.1016/j.chemolab.2017.02.008.

[8] E. Szymańska, Modern data science for analytical chemical data – A comprehensive review, Anal. Chim. Acta. 1028 (2018) 1–10. https://doi.org/10.1016/j.aca.2018.05.038.

[9] V. Di Egidio, N. Sinelli, G. Giovanelli, A. Moles, E. Casiraghi, NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, Eur. Food Res. Technol. 230 (2010) 947–955. https://doi.org/10.1007/s00217-010-1227-5.

[10] S. Buratti, D. Ballabio, G. Giovanelli, C.M. Zuluanga-Dominguez, A. Moles, S. Benedetti, N. Sinelli, Monitoring of alcoholic fermentation using near infrared and mid infrared spectroscopies combined with electronic nose and electronic tongue, Anal. Chim. Acta. 697 (2011) 67–74. https://doi.org/10.1016/j.aca.2011.04.020.

[11] M. Malfeito-Ferreira, Wines: Wine Spoilage Yeasts and Bacteria, Second Edi, Elsevier, 2014. https://doi.org/10.1016/B978-0-12-384730-0.00390-6.

[12] Cavaglia, J., Schorn-García, D., Giussani, B., Ferré, J., Busto, O., Aceña, L., Mestres, M., & Boqué, R. (2020). ATR-MIR spectroscopy and multivariate analysis in alcoholic fermentation monitoring and lactic acid bacteria spoilage detection. Food Control, 109. https://doi.org//10.1016/j.foodcont.2019.106947

[13] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data, Bioinformatics. 21 (2005) 3043–3048. https://doi.org/10.1093/bioinformatics/bti476.

[14] S. Grassi, I. Vigentini, N. Sinelli, R. Foschino, E. Casiraghi, Near Infrared and Mid Infrared Spectroscopy in Oenology: Determination of Main Components Involved in Malolactic Transformation, NIR News. 23 (2012) 11–14. https://doi.org/10.1255/nirn.1300.

[15] J. Manuel Amigo, A. del Olmo, M. Møller Engelsen, H. Lundkvist, S. Balling Engelsen, Staling of white wheat bread crumb and effect of maltogenic α-amylases. Part 2: Monitoring the staling process by using near infrared spectroscopy and chemometrics, (2019). https://doi.org/10.1016/j.foodchem.2019.06.013.

[16] S. Grassi, C.B. Lyndgaard, M.A. Rasmussen, J.M. Amigo, Interval ANOVA simultaneous component analysis (i-ASCA) applied to spectroscopic data to study the effect of fundamental fermentation variables in beer fermentation metabolites, Chemom. Intell. Lab. Syst. 163 (2017) 86–93. https://doi.org/10.1016/j.chemolab.2017.02.010.

Article 3

"ATR-MIR spectroscopy and multivariate analysis in alcoholic fermentation monitoring and lactic acid bacteria spoilage detection"

Published in Food Control, Volume 109, March 2020, 106947

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

ATR-MIR spectroscopy and multivariate analysis in alcoholic fermentation monitoring and lactic acid bacteria spoilage detection

Julieta Cavaglia[a] | Daniel Schorn-García[a] | Barbara Giussani[b] | Joan Ferré[c] | Olga Busto[a] | Laura Aceña[a] | Montserrat Mestres[a] | Ricard Boqué[c]*

[a.] Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

[b.] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio, 9, Como (22100), Italy

[c.] Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

*Corresponding author: ricard.boque@urv.cat

Abstract

Wine production processes still rely on post-production evaluation and off-site laboratory analyses to ensure the quality of the final product. Here we propose an *at-line* methodology that combines a portable ATR-MIR spectrometer and multivariate analysis to control the alcoholic fermentation process and to detect wine fermentation problems. In total, 36 microvinifications were conducted, 14 in *normal fermentation conditions* (NFC) and 22 *intentionally contaminated fermentations* (ICF) with different lactic acid bacteria (LAB) concentrations. ATR-MIR measurements were collected during alcoholic and malolactic fermentations and relative density, pH, and L-malic acid were analyzed by traditional methods. Partial Least Squares Regression could suitably predict density and pH in fermenting samples (root mean squared errors of prediction of 0.0014 $g \cdot mL^{-1}$ and 0.06 respectively). With regard to ICF, LAB contamination was detected by multivariate discriminant analysis when the difference in L-malic acid concentration between NFC and ICF was in the order of 0.7-0.8 $g \cdot L^{-1}$, before the end of malolactic fermentation. This methodology shows great potential as a fast and simple *at-line* analysis tool for detecting fermentation problems at an early stage.

Keywords: Process monitoring, alcoholic fermentation, wine, malolactic fermentation contamination, ATR-MIR, Process Analytical Technologies

Acknowledgments: This work was supported by the Spanish Ministry of Science and Technology and the European Union (MINECO-FEDER) in the Project AGL2015-70106-R and the Catalan Research Council (AGAUR) for the FI Grant 2019 awarded to Cavaglia, J. (Record Number FI_B100154).

1. Introduction

The production of wine is based on alcoholic fermentation, which consists in the biochemical transformation of sugar into ethanol by yeasts. There are many factors that have an influence over the complexity and quality of the final product such as the grape quality and variety, yeast strain or cellar practices used (Suárez-Lepe & Morata, 2012). However, even with the best raw materials and starting under the optimal conditions, problems during alcoholic fermentation can occur, in which yeast or other microorganisms synthetize undesirable compounds that negatively affect the quality of the wine. Stuck and sluggish fermentations along with contamination-related processes are the most common problems that can appear during alcoholic fermentation (Hernández, León, & Urtubia, 2016). Nutrient deficiencies, sudden temperature changes or the imposition of undesired and non-inoculated yeast are the main causes of stuck and sluggish fermentations. Spoilage processes are due to the growth of unwanted microorganisms in the must, such as acetic acid or lactic acid bacteria (LAB), which are part of the normal microbiota found on the surface of leaves and grapes but can also be found in the environment of wineries (Portillo, Franquès, Araque, Reguant, & Bordons, 2016). Although the "piqûre acétique" is the most widely known spoilage, the "piqûre lactique" can also pose very important problems in some wines.

LAB are responsible for the biochemical transformation of L-malic acid into L-lactic acid releasing carbon dioxide. This process, called malolactic fermentation, is promoted in red wines to decrease their acidity since, from an organoleptic point of view, a lower acidity is more compatible with the high tannicity of these wines (Cappello, Zapparoli, Logrieco, & Bartowsky, 2017). However, in white wines, this second fermentation is usually undesired because it increases pH and reduces their typical freshness, leading to wines with worse organoleptic quality (Cozzolino, Mccarthy, & Bartowsky, 2012).

In the winemaking industry, a control of the alcoholic fermentation process is required in order to avoid problems that result in low quality wines and consequently, in economic losses. In the cellar, the process is mostly controlled by determining temperature, density and pH, which are usually measured twice a day, together with a visual and aroma evaluation of the fermenting grape must. These parameters are related to sugars, acids and other minor compounds that ultimately impact substantially the colour and/or aroma of the wine (Bisson, 1999). These parameters are sufficient to control the process when the fermentation progresses well. However, these control measures sometimes fail to timely detect problems when they could still be solvable by applying corrective measures to the

must. This is why the implementation of novel process control strategies to obtain real-time information during alcoholic fermentation has a growing interest in the oenological field (Cozzolino, 2016).

The Process Analytical Technologies (PAT) approach follows this trend. PAT is a system for designing, analysing and controlling a manufacturing process, through timely measurements of critical quality attributes of raw and in-process materials and processes in order to ensure final product quality. The hypothesis behind PAT is that quality must be controlled through process control and not only by evaluating postproduction information (Simon, Pataki, Marosi, Meemken, Hungerbühler, *et al.*, 2015). This is especially advantageous when applied over expensive or complex samples such as pharmaceuticals or food products (Lourenço, Lopes, Almeida, Sarraguça, & Pinheiro, 2012; Van Den Berg, Lyndgaard, Sørensen, & Engelsen, 2013). For this reason, the winemaking industry is a sector where PAT could be widely applied.

In the last decades, the use of spectroscopy to determine oenological parameters has increased considerably. Spectroscopic methods are fast, clean and provide large amounts of information with minimum sample preparation. Near and Mid Infrared Spectroscopy (FT-NIR and FT-MIR) have been widely used to monitor wine fermentations because information can be obtained on-time all along the process (Buratti, Ballabio, Giovanelli, Zuluanga Dominguez, Moles *et al.*, 2011; Urtubia, Pérez-Correa, Meurens, & Agosin, 2004). Several authors have reported good prediction of sugars (glucose and fructose), ethanol, volatile acids, phenolic compounds or volumic mass in must, fermenting must and wine samples (Cozzolino, 2016; Di Egidio, Sinelli, Giovanelli, Moles, & Casiraghi, 2010; dos Santos, Páscoa, & Lopes, 2017). In some cases, the prediction of chemical parameters has allowed detecting some problems such as sluggish fermentations (Urtubia, Pérez-Correa, Pizarro, & Agosin, 2008). Among these studies, those using MIR spectroscopy with attenuated total reflectance (ATR-MIR) stand out because this technique only requires one drop of sample and provides well resolved water peaks (Teixeira dos Santos, Páscoa, & Lopes, 2017; Shah, Cynkar, Smith, & Cozzolino, 2010). All these advantages, together with the fact that modern MIR spectrometers can also be portable, make this technique a very suitable tool in a cellar not only to monitor different fermentation parameters but also to detect fermentation problems as we demonstrated in a previous study (Cavaglia, Giussani, Mestres, Puxeu, Busto, *et al.*, 2019).

The present research aims to evaluate the application of a portable ATR-MIR spectrometer and multivariate analysis techniques to control the progress of alcoholic fermentations and

to detect problems at an early stage. Density and pH were evaluated using regression models, whereas discriminant models were used to detect wine fermentation deviations due to LAB contamination.

2. Materials and methods

2.1. Grape must and microorganisms

Concentrated white must was provided by Mostos Españoles S.A., (Ciudad Real, Spain) and it was stored at -20 ºC until its use. Its defrosting was done at 5 ºC and it was then diluted with MilliQ water to adjust the sugar concentration to 200 ± 10 $g \cdot L^{-1}$. The diluted must was supplemented with 0.30 $g \cdot L^{-1}$ of ENOVIT® (SPINDAL S.A.R.L. Gretz Armainvilliers, France) and 0.30 $g \cdot L^{-1}$ of Actimaxbio* (Agrovin, Ciudad Real, Spain) in order to ensure a sufficient final concentration of yeast assimilable nitrogen.

The commercial dry *Saccharomyces cerevisiae* yeast strain used was "E491" (Vitilevure Albaflor, YSEO, Danstar Ferment A.G., Denmark). Regarding to lactic acid bacteria, a commercial freeze-dried blend of *Oenococcus oeni* and *Lactobacillus plantarum* "Co-inoculant Bacteria 3.2" (Anchor Oenology, South Africa) was used. Rehydration of the microorganisms was done following the suppliers' instructions.

2.2. Microvinifications

Three small-scale alcoholic fermentation or microvinification batches were carried out as follows. For each sample, 350 mL of diluted must were added into 500 mL Erlenmeyer flasks and they were inoculated with 0.105 g of active dry yeast rehydrated in 2 mL of MiliQ water for 30 minutes at 25 ºC, reaching a final concentration of $3 \cdot 10^6$ $CFU \cdot mL^{-1}$. To prepare the simulated contaminated samples, LAB co-inoculations were done taking into account the producer instructions (1 g = $1 \cdot 10^{11}$ $CFU \cdot mL^{-1}$) to reach different final concentrations ranging between 1·106 and $1 \cdot 10^7$ $CFU \cdot mL^{-1}$. All microvinifications were kept under a constant temperature of 18ºC until the end of alcoholic and malolactic fermentations. Alcoholic fermentation was considered finished when density was under 0.995 $g \cdot L^{-1}$ whereas malolactic fermentation ended when L-malic acid concentration was < LOD (0.06 $g \cdot L^{-1}$).

The number of samples of each batch, the initial must parameter values (which are slightly different to simulate the natural maturity variability in grapes) and codification used are specified in Table 1. The normal fermentation conditions were coded as NFC and the intentionally contaminated fermentations as ICF. ICF samples were divided into 5 groups:

ICF1, ICF2, ICF3, ICF4 and ICF5, according to the concentrations of LAB inoculated. The aim of using different concentrations of LAB was to promote the transformation of L-malic acid into L-lactic acid at different points of the alcoholic fermentation.

Table 1. Initial conditions for the 3 experiments conducted. Yeast concentration was $3x10^6$ CFU·mL^{-1} in all batches.

	Number of samples	Sample name	Inoculated LAB population (CFU·mL^{-1})	Initial density (g·mL^{-1})	Initial pH	Initial L-Malic Acid (g·L^{-1})
Batch 1	5	NFC	-	1.0872	4.15	1.92
	5	ICF1	$1·10^6$			
Batch 2	3	NFC	-	1.0830	4.09	1.73
	3	ICF1	$1·10^6$			
	3	ICF2	$1.6·10^6$			
	3	ICF3	$2.5·10^6$			
Batch 3	6	NFC	-	1.0794	4.04	1.61
	4	ICF4	$7·10^6$			
	4	ICF5	$1·10^7$			

2.3. ATR-MIR analysis

The samples were collected at least once a day to follow both alcoholic and malolactic fermentations until both were finished. The sampled volume was 1.5 mL, which was centrifuged at 10000 rpm for 10 minutes to avoid the scattering effect in the spectroscopic measurements due to the presence of microorganisms. The pellet was discarded while the supernatant was kept in 1.5 mL eppendorfs for further analysis. Right after sample collection, 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. A drop of sample was placed onto the crystal using a Pasteur pipette and the spectra were acquired right afterwards. Each sample was analysed in triplicate and an air background was recorded between samples. Each sample was measured applying our previously optimized methodology (Cavaglia, Giussani, Mestres, Puxeu, Busto, *et al.*, 2019). After each measurement, the crystal was thoroughly cleaned with deionized water and cotton wipes. Spectra were collected in absorbance mode from 4000 to 650 cm^{-1}. The resolution and number of scans that provided the best results were 8 cm^{-1} and 32, respectively. Measurements were made at 63 ± 1 ºC, as this was the

stabilization temperature of the crystal. Spectra were examined using the Microlab PC software (Agilent, California, USA), and saved as .spc files.

2.4. Standard sample analysis

As it is done in a cellar, density and pH were determined along the alcoholic fermentation to ensure the normal progress of this process. Density was measured using an electronic portable densimeter (Densito2Go, Mettler Toledo, United States) and pH was measured with a portable pH-meter with a 201 T portable electrode (7+ series portable pH-meter, XS Instruments, Italy). The remaining volume of the supernatant was used for L-malic acid analysis using an Y15 Analyser (Biosystems, Barcelona, Spain) in order to follow the malolactic fermentation. Measurements were performed in parallel to ATR-MIR analysis of the samples.

2.5. Multivariate analysis

For each sample, the average of the three recorded spectra was used in all the models described below.

The collected data had a three-dimensional structure, with I samples, J wavenumbers and K sampling times. This 3-way array was rearranged in different ways (Figure 1), depending on the aim of the study.

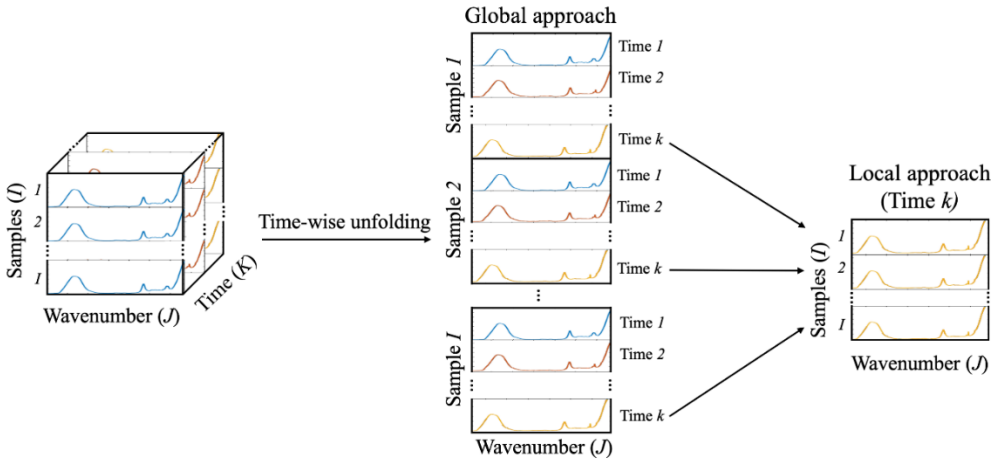


Figure 1. Scheme of the data arrangement for each of the approaches applied to the three-way structure.

First, a global approach was developed using all the spectra collected throughout the fermentation process for all the experiments, to explore the main information contained in the data and to correlate the spectra with the fermentation parameters. A time-wise unfolding of the 3-way array was performed to obtain a matrix with dimensions (IK x J), in

which rows were the spectra recorded for I samples at K sampling times and columns were the J spectroscopic wavenumbers.

The individual examination of each sampling time was also considered for the discrimination between NFC and ICF samples for each experimental batch following a local approach. For each experimental batch, different K matrices (one for each time sampled) with dimensions (I x J) were thus independently investigated.

Principal Component Analysis (PCA) was applied to visualize the variability among data both through alcoholic and malolactic fermentations and to detect potential outliers, while Partial Least Squares Regression (PLSR) models were developed to predict fermentation parameters.

Finally, Partial Least Squares Discriminant Analysis (PLS-DA) was used to detect LAB spoilage. PLS-DA is similar to PLSR, but in this method the vector y contains dummy variables (0 or 1) for the classes you want to discriminate (here, NFC and ICF). The method seeks the optimal number of latent variables (LVs) that maximize the covariance (and thus the discrimination) between the infrared spectra and the classes. A discrimination threshold (between 0 and 1) is calculated taking into account the probability of classification error of the samples into the classes (Pérez, Ferré, & Boqué, 2008).

To proceed with the study of the spectra, different pre-processing strategies were tested including first and second derivatives (to emphasise small peaks), Savitzky-Golay smoothing (to reduce noise) and Standard Normal Variate (SNV) (to reduce the variability between samples due to scatter). This step is crucial because the outcome of a multivariate model has a strong dependence on the pre-processing applied. According to the data matrix used in the calculation, different pre-processing combinations were tried and compared. Only those giving the best results are shown. After spectral pre-processing, data were mean-centered. The theoretical basis of these treatments can be found elsewhere (Rinnan, Van Den Berg, & Engelsen, 2009).

In addition, to optimize the regression models and further reduce their complexity, a variable selection strategy based on the Selectivity Ratio algorithm was considered. It is based on the idea of progressively excluding variables in the X data block and evaluate the effectiveness of the Y prediction until the combination of X variables is optimized (Rajalahti, Arneberg, Berve, Myhr, Ulvik, *et al.*, 2009).

Regression models were validated considering three different validation strategies and the best model was selected by evaluating the best compromise between the higher percentage

of explained variance in Y and the minimum RMSECV/RMSEP (Root Mean Square Error of Cross-Validation/Prediction). In the first validation strategy, an internal cross-validation (CV) was performed, where groups of samples (accounting for 5% of the total number of samples) were left out each time and used for prediction. The procedure was iterated 20 times and the average RMSECV was considered. In the second strategy, data were split into random halves and each half was used as calibration set in one model and as validation test set in the other. Thus, a random vector of zeros and ones was built, where zeros were considered calibration samples and ones were validation samples. The data split procedure was repeated 10 times to reduce the dependence of data splits in the performance of the models and the average RMSEP error was evaluated. Finally, the third strategy consisted on applying the Kennard-Stone sample selection algorithm which divides the data into calibration and test sets taking into account the distribution of the samples in the principal components space. This algorithm selects the samples for the calibration set providing uniform coverage over the X data, including samples at the limits of the measurements ranges (Kennard & Stone, 1969). This methodology tends to be overoptimistic, and for this reason the number of samples to be included in the calibration test was optimized, assuring a RMSEP comparable to the ones obtained by the other strategies.

In the case of the PLS-DA models, different internal CV strategies were tested, depending on the number of samples available in each case. A leave-one-out CV was used when the number of samples ≤6, while a leave-two-out CV was used when the number of samples ≥6.

All multivariate data analyses were performed using the PLS Toolbox v8.7 (Eigenvector Research Inc., Eaglerock, USA) with MATLAB R2015b (The MathWorks, Natick, USA).

3. Results and discussion

3.1. ATR-MIR spectra

The evolution of the ATR-MIR spectra during alcoholic fermentation is shown in Figure 2. The region from 850 to 649 cm^{-1} was excluded as it did not contain useful information, resulting in low quality models. As previously reported (Cozzolino & Curtin, 2012; Wu, Xu, Long, Zhang, Wang, *et al.*, 2015), the regions that show most of the variability during wine alcoholic fermentation in the mid-infrared region are mainly found between 950 to 1500 cm^{-1}, where CH2, C-C-H, H-C-O bonds and C-C, C-O stretching vibrations absorb, and between 3000 to 3700 cm^{-1}, where O-H stretching absorbs.

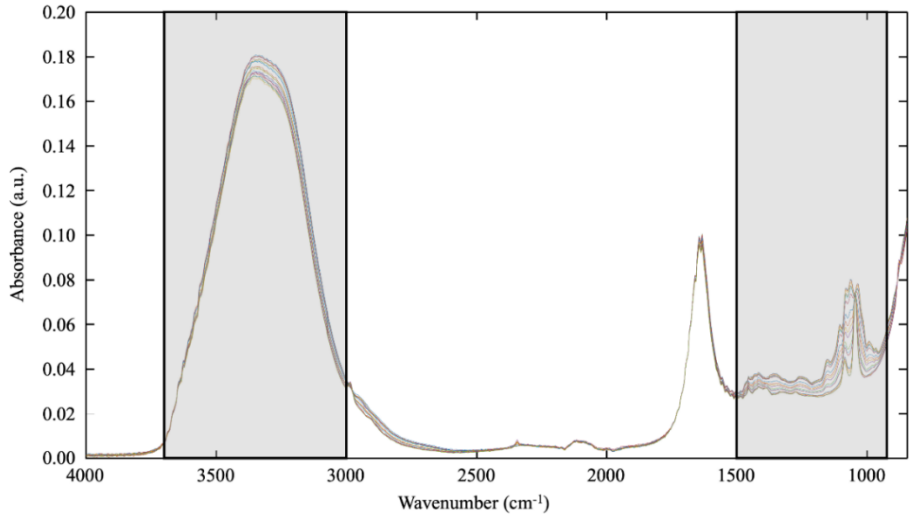


Figure 2. Evolution of the full raw spectra in ATR-MIR mode highlighting in grey the areas that show most of the variability during alcoholic fermentation.

3.2. Alcoholic fermentation

All the spectra arranged in a time-wise unfolded matrix (Figure 1) were used to build a global PCA model. The evolution of each batch during alcoholic fermentation was best described when applying the following pre-processing combination: Savitzky-Golay second order polynomial smoothing through 7 points, SNV and mean-centering. The first 2 principal components accounted for the 99.31% of the data variability (97.39% for PC1 and 1.92% for PC2). As it can be noticed in Figure 3, when comparing the evolution in time of the PC1 scores with the evolution of the density curve with the values registered during the fermentation process, both plots show a similar trend. The loadings plot of PC1 shows that the most important region to follow the progress of alcoholic fermentation is between 950-1700 cm^{-1} (data not shown), which was not surprising as this region mainly corresponds to sugars and ethanol absorptions (Cozzolino, Cynkar, Shah & Smith, 2011).

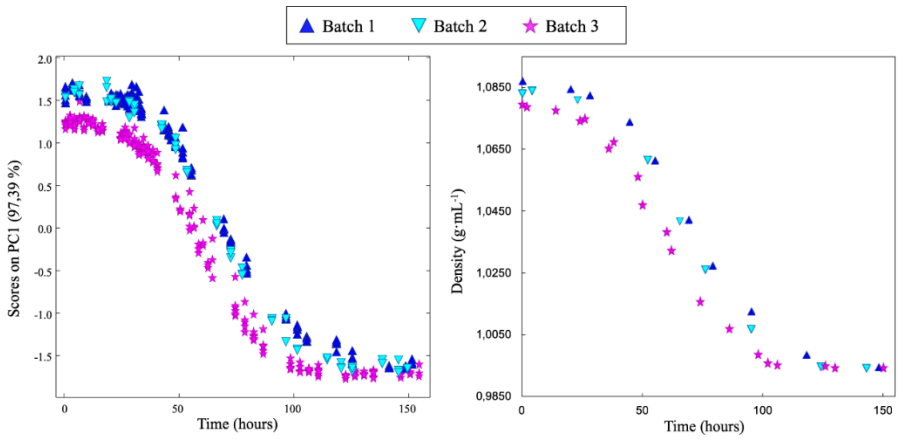


Figure 3. PC1 score values of the global model for all times (left). Density trend during alcoholic fermentation. Samples are labelled according to the batch experiment (right).

Moreover, the PCA model built using the spectra shows small differences between batches. A hypothesis is that this behaviour could be related to small changes in the initial sample density, since all samples come from the dilution of the same must in the same experimental conditions. In other words, the spectra recorded by the portable instrument allowed to distinguish between experiments, confirming the capability of the spectroscopic technique coupled with chemometrics to spot small differences between fermentation processes.

3.3. Prediction of chemical parameters

As mentioned above, PC1 scores and density showed a similar trend when depicted against time. From this important result arose the idea of using the spectroscopic data to predict density by means of PLSR. All the available NFC experiments were used in this regression model (final data matrix dimensions 566 samples x 850 variables).

By applying the Selectivity Ratio algorithm, the spectroscopic regions selected were 967 to 1175 cm^{-1} and 1483 to 1771 cm^{-1}. The validation errors for the density models using the different CV strategies are shown in Table 2. For the first model, a subset of 28 samples was used. The number of LVs to be considered was optimized taking into account the higher percentage of explained variance of Y data and the lower RMSECV/ RMSEP values. For the subsequent models, only one LV was used. The Kennard-Stone algorithm showed that only 29 calibration samples were necessary to build a model with an RMSEP value comparable to the ones obtained by the other validation methods.

Table 2. Results of the three different methods of validation, shown as RMSECV for internal CV, and RMSEP for halves splitting and Kennard Stone validation for density and pH prediction.

	Density ($g \cdot mL^{-1}$)	pH
Internal CV	0.0014	0.06
Halves splitting	0.0014	0.07
Kennard Stone	0.0014	0.06

Similar results have been reported using NIR spectroscopy. Fernandez-Novales *et al.* obtained an RMSECV of 0.0065 $g \cdot mL^{-1}$ for the prediction of density in wine fermenting samples (Fernández-Novales, López, González-Caballero, Ramírez, & Sánchez, 2011). In our study, we showed for the first time that the spectroscopic information obtained with a portable ATR-MIR spectrometer with PLSR can be used to predict density in must and fermenting samples, obtaining very satisfactory results considering the lower optical robustness of the instrument compared to benchtop devices.

pH is another chemical parameter that is usually determined to control alcoholic fermentation. In this study, PLSR was applied to predict pH following the same methodology as for density (in this case, the data matrix dimensions were 427 samples x 850 variables). The Selectivity Ratio algorithm selected regions all along the spectroscopic range, suggesting that pH prediction requires information from the full spectrum. A combination of Savitzky-Golay second order polynomial smoothing through 15 points, SNV and mean-centering pre-processing gave the best results. For all models, 5 LVs were needed to achieve good predictions. In the first model built with all the samples, a subset of 22 samples was used for internal validation. The validation based on the Kennard-Stone selection method needed 43 calibration samples to obtain errors comparable to those of the other validation methods; therefore, 384 validation samples were used to test the model. Results from the different validation strategies for the pH models are summarised in Table 2. Swanepoel *et al.* obtained a standard error of prediction (SEP) of 0.05 pH units for grape and must samples using FT-MIR in the transmission mode (Swanepoel, du Toit, & Nieuwoudt, 2007). Using ATR-MIR, Shah *et al.* obtained a standard error of cross-validation (SECV) of 0.07 for the pH of grape juice samples (Shah, Cynkar, Smith, & Cozzolino, 2010). Our results show that the portable spectrometer used in this study can perform a fast and simple control of the progress of alcoholic fermentation with an acceptable error when combined with a chemometric strategy to manage the recorded

spectra. Additionally, the fact that similar validation errors were obtained using different validation strategies shows the robustness of the models.

3.4. Malolactic fermentation deviation

The spectra recorded during the experiments in which LAB co-inoculations were performed (ICF) showed only minor changes with respect to the ones recorded in NFC due to the small concentration changes involved in the malolactic fermentation process. The main information in both NFC and ICF spectra is ascribable, in fact, to the alcoholic fermentation evolution (sugars and ethanol signals).

To focus the attention on the malolactic fermentation process, each batch was individually studied to avoid the variability among batches. In addition, the PCA models were built using the spectroscopic region between 1320 and 1109 cm^{-1}, which is related to organic acid molecules involved in the malolactic fermentation as previously reported (Grassi, Vigentini, Sinelli, Foschino, & Casiraghi, 2012; Picque, Lefier, Grappin, & Corrieu, 1993).

Three models were calculated, one for each batch experiment. The best results were obtained with a combination of 1st derivative Savitzky-Golay second order polynomial smoothing through 15 points, SNV and mean-centering as pre-processing methods. In this case, the 1st derivative emphasised the slight changes in small peaks. All models explained more than 98% of the variability using 3 PCs.

The scores for two PCs against time are depicted for each batch in Figure 4. Samples are labelled according to the LAB co-inoculated concentrations. It can be observed that the evolution of malolactic fermentation takes different directions in the PCA space with respect to time and it is even possible to distinguish among the different LAB concentrations in the second and third batches. The models allowed to observe the different trends between ICF and NFC samples before the end of malolactic fermentations, and in some cases, before the end of alcoholic fermentation (Batch 3). A deep investigation of these plots allowed to qualitatively determine at which sampling time the trajectories of ICF samples started to deviate from NFC. In batches 1 and 2, trajectories showed different trends 100 hours after the beginning of alcoholic fermentation, whereas in batch 3, it was possible to qualitative see the different trajectories after 50 hours.

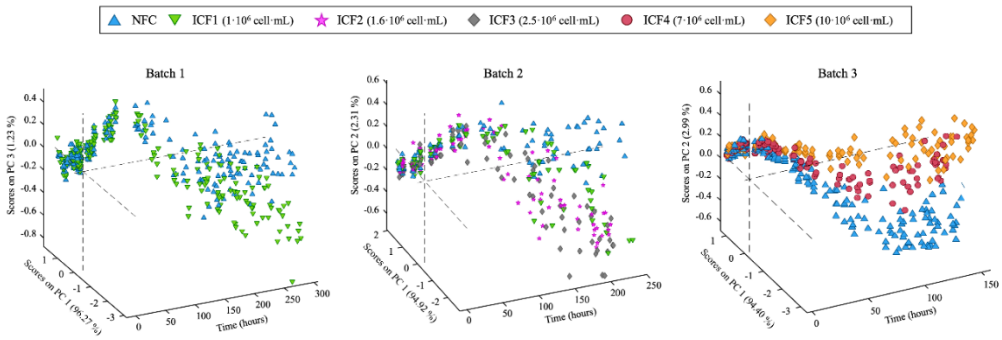


Figure 4. Score plots of the PCA global model for each batch using only the selected spectroscopic region.

3.5. Discrimination between NFC and ICF

Starting from the qualitative results previously shown (section 3.4), PLS-DA models for each batch were built at individual sampling times (local models) to determine at which sampling time the trajectories of ICF samples started to deviate from NFC. In other words, to determine as soon as possible when the deviation from the NFC occurred because of LAB spoilage. For each PLS-DA model at each sampling time (Figure 1) the y vector was built by assigning 1s to ICF samples and 0s to NFC samples.

The first sampling time to find a discrimination threshold between the two groups with a 100% correct classification was defined as the deviation time. The deviation time was confirmed with a local model of the consecutive sampling time when 100% correct classification was achieved. For all models, only one LV was needed for a successful discrimination of the classes.

Samples deviated from NFC in the first batch 213 hours after the beginning of the fermentation. In the second batch, ICF1 deviated after 187 hours whereas ICF2 and ICF3 deviated after 145 hours and 138 hours, respectively. In the third batch the difference of ICF4 and ICF5 from NFC was detected after 56 and 58 hours, respectively.

At those deviation times, malolactic fermentation was around 50%-60%, which means that it is possible to differentiate the spectra before the end of malolactic fermentation, allowing to make corrective measures in wineries. Manley *et al.* considered the possibility of using FT-NIR to detect if malolactic fermentation has started, is in progress or has been completed in white wine, where L-lactic acid values were between 0-0.3 $g \cdot L^{-1}$, 0.3-2 $g \cdot L^{-1}$ and above 3 $g \cdot L^{-1}$, respectively. They reported good classification of each class, with >95% of recognition rates (Manley, van Zyl, & Wolf, 2001). In our study, for all PLS-DA models,

the difference in L-malic acid concentration between NFC and ICF samples ranged from 0.7 to 0.8 $g \cdot L^{-1}$. Despite the fact that this decrease in L-malic acid concentrations result in a slight increase in pH, this is the first time that an ATR-MIR device is used to detect deviations from NFC before the end of malolactic fermentation.

4. Conclusions

It has been demonstrated that a portable ATR-MIR spectrometer with multivariate analysis is a valuable analytical tool to rapidly control the progress of alcoholic fermentation in white wine. Here, the ability of this portable device has been proved to effectively predict density and pH in fermenting must samples. The methodology presented shows great potential as a fast and simple at-line analysis tool for the detection of fermentation problems, as is possible to use this instrument to rapidly assess a LAB spoilage during alcoholic fermentation. Upon this findings, further research will be developed based on PAT strategies to give the winemaker the possibility to correct the process and to obtain good quality wines.

5. References

Bisson, L. F. (1999). Stuck and sluggish fermentations. American Journal of Enology and Viticulture, 50(1), 107–119.

Buratti, S., Ballabio, D., Giovanelli, G., Zuluanga Dominguez, C. M., 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. https://doi.org/10.1016/j.aca.2011.04.020.

Cappello, M. S., Zapparoli, G., Logrieco, A., & Bartowsky, E. J. (2017). Linking wine lactic acid bacteria diversity with wine aroma and flavour. International Journal of Food Microbiology, 243, 16–27. https://doi.org/10.1016/j.ijfoodmicro.2016.11.025.

Cavaglia, J., Giussani, B., Mestres, M., Puxeu, M., Busto, O., Ferré, J., & Boqué, R. (2019). Early detection of undesirable deviations in must fermentation using a portable FTIR-ATR instrument and multivariate analysis. Journal of Chemometrics, (e3162), 1–11. https://doi.org/. https://doi.org/10.1002/cem.3162.

Cozzolino, D., & Curtin, C. (2012). The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments. Food Control, 26, 241–246. https://doi.org/10.1016/j.foodcont.2012.02.006.

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(4), 282–297. https://doi.org/10.1080/05704928.2015.1132721.

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. https://doi.org/10.1016/j.foodcont.2012.01.003.

Cozzolino D, Cynkar W, Shah N & Smith P. (2011) Feasibility study on the use of attenuated total reflectance mid-infrared for analysis of compositional parameters in wine. Food Research International, 44(1), 181–186.

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. https://doi.org/10.1007/s00217-010-1227-5.

Fernández-Novales, J., López, M. I., González-Caballero, V., Ramírez, P., & Sánchez, M. T. (2011). Feasibility of using a miniature NIR spectrometer to measure volumic mass during alcoholic fermentation. International Journal of Food Sciences and Nutrition, 62(4), 353–359. https://doi.org/10.3109/09637486.2010.533161.

Grassi, S., Vigentini, I., Sinelli, N., Foschino, R., & Casiraghi, E. (2012). Near Infrared and Mid Infrared Spectroscopy in Oenology: Determination of Main Components Involved in Malolactic Transformation. NIR News, 23(3), 11–14. https://doi.org/10.1255/nirn.1300.

Hernández, G., León, R., & Urtubia, A. (2016). Detection of abnormal processes of wine fermentation by support vector machines. Cluster Computing, 19, 1219–1225. https://doi.org/10.1007/s10586-016-0594-5.

Kennard, R. W., & Stone, L. A. (1969). Technometrics Computer Aided Design of Experiments. Technometric, 11(1), 137–148. http://dx.doi.org/10.1080/00401706.1969.10490666.

Lourenço, N. D., Lopes, J. A., Almeida, C. F., Sarraguça, M. C., & Pinheiro, H. M. (2012). Bioreactor monitoring with spectroscopy and chemometrics: A review. Analytical and Bioanalytical Chemistry, 404(4), 1211–1237. https://doi.org/10.1007/s00216-012-6073-9.

Manley, M., van Zyl, A., & Wolf, E. E. H. (2001). The Evaluation of the Applicability of Fourier Transform Near-Infrared (FT-NIR) Spectroscopy in the Measurement of

Analytical Parameters in Must and Wine. South African Journal of Enology & Viticulture, 22(2), 93–100. https://doi.org/https://doi.org/10.21548/22-2-2201.

Pérez, N. F., Ferré, J., & Boqué, R. (2009). Calculation of the reliability of classification in discriminant partial least-squares binary classification. Chemometrics and Intelligent Laboratory Systems, 95, 122–128. https://doi.org/10.1016/j.chemolab.2008.09.005.

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. https://doi.org/10.1016/0003-2670(93)85067-T.

Portillo, M. del C., Franquès, J., Araque, I., Reguant, C., & Bordons, A. (2016). Bacterial diversity of Grenache and Carignan grape surface from different vineyards at Priorat wine region (Catalonia, Spain). International Journal of Food Microbiology, 219, 56–63. https://doi.org/10.1016/j.ijfoodmicro.2015.12.002.

Rajalahti, T., Arneberg, R., Berven, F. S., Myhr, K.-M., Ulkiv, R. J. & Kvalheim, O. M. (2009). Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. Chemometrics and Intelligent Laboratory Systems, 95, 35–48. https://doi.org/10.1016/j.chemolab.2008.08.004

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. https://doi.org/10.1021/jf100420z.

Simon, L. L., Pataki, H., Marosi, G., Meemken, F., Hungerbühler, K., Baiker, et. al. (2015). Assessment of recent process analytical technology (PAT) trends: A multiauthor review. Organic Process Research and Development, 19(1), 3–62. https://doi.org/10.1021/op500261y.

Suárez-Lepe, J. A., & Morata, A. (2012). New trends in yeast selection for winemaking. Trends in Food Science & Technology, 23, 39–50. https://doi.org/10.1016/j.tifs.2011.08.005.

Swanepoel, M., du Toit, M., & Nieuwoudt, H. H. (2007). Optimisation of the quantification of total soluble solids, pH and titratable acidity in South African grape must using fourier transform mid-infrared spectroscopy. South African Journal of Enology and Viticulture, 28(2), 140–149.

Teixeira dos Santos, C.A., 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. https://doi.org/10.1016/j.trac.2016.12.012.

Urtubia, A., Pérez-Correa, J. R., Meurens, M., & Agosin, E. (2004). Monitoring large scale wine fermentations with infrared spectroscopy. Talanta, 64, 778–784. https://doi.org/10.1016/j.talanta.2004.04.005.

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. https://doi.org/10.1016/j.foodcont.2007.04.017.

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. https://doi.org/10.1016/j.tifs.2012.04.007.

Wu, Z., Xu, E., Long, J., Zhang, Y., Wang, F., Xu, X., et. al. (2015). Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy. Food Control, 50, 405–412. https://doi.org/10.1016/j.foodcont.2014.09.028.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chapter 5

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Article 4

"Monitoring wine fermentation deviations using an ATR-MIR spectrometer and MSPC charts"

Published in Chemometrics and Intelligent Laboratory Systems
Volume 201, 15 June 2020, 104011

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

This chapter focus on the development of robust MSPC charts based on the PCA models of experimental data from microvinification studies.

The aim of this study was to apply different MSPC strategies to monitor wine alcoholic fermentation. The last specific objective of the thesis was addressed, as the control charts were developed to detect deviations based on the modelling of fermentations running under Normal Operation Conditions (NOC). These charts can then be used to monitor and eventually control the correct progression of wine alcoholic fermentation. The idea behind this PAT methodology is that, under NOC conditions, PCA on spectroscopic data can be used to describe a NOC space and then the corresponding Hotelling T^2 and Q statistical limits (i.e. at 95% confidence) can be calculated. Then, any sample that is projected onto the NOC-PCA defined space will have a value for the two statistics. If a sample is deviated from the process, it is expected that at some time it will show unusual T^2 or Q values (out-of-control values).

In addition, as ATR-MIR spectra change considerably during alcoholic fermentation, eight-hour interval Q-charts were developed to improve the performance of the process control monitoring.

The variability between different batches and experiments was studied, as process control strategies must consider this variability factor in order to develop valid models.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Monitoring wine fermentation deviations using an ATR-MIR spectrometer and MSPC charts

Julieta Cavaglia[a] | Daniel Schorn-García[a] | Barbara Giussani[b] | Joan Ferré[c] | Olga Busto[a] | Laura Aceña[a] | Montserrat Mestres[a] | Ricard Boqué[c]*

[a.] Universitat Rovira i Virgili. Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

[b.] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio, 9, Como (22100), Italy

[c.] Universitat Rovira i Virgili. Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain

*Corresponding author: ricard.boque@urv.cat

Abstract

Despite the winemaker's efforts, deviations such as bacterial spoilage can occur during wine alcoholic fermentation resulting in economic losses and low quality wines. When a deviation is suspected, samples are usually sent to an oenological laboratory for the off-line analysis of specific quality control parameters. The use of ATR-MIR as a fast analytical tool to monitor the fermentation process could be very useful, as getting real-time information of the process allows making readjustments before the process ends. In this study, we aimed at detecting white wine spoilage during alcoholic fermentation due to the action of lactic bacteria using a portable ATR-MIR instrument and MSPC charts. A total of 33 small-scale alcoholic fermentations were conducted (25 in normal operation conditions (NOC) and 8 simulating a bacterial spoilage with the addition of lactic bacteria (MLF)) to evaluate the capability of the MSPC charts to detect deviations from NOC. MSPC control charts were developed based on Q residuals and Hotelling's T^2 statistics. Time-wise unfolding was applied to the original three-way data to build different PCA models, obtaining very satisfactory results: MLF samples were detected before the end of alcoholic fermentation in the Q residuals charts after 80 hours and Hotelling T^2 chart could also differentiate the samples after 100 hours.

Key words: Wine Alcoholic fermentation, ATR-MIR, MSPC, Process analytical technologies (PAT), Quality control.

1. Introduction

The main biochemical change during wine alcoholic fermentation is the transformation of sugars from grape must into ethanol by the action of yeasts. In order to obtain high quality wines a close monitoring of this process is of utmost importance [1]. In the wine cellar, simple measurements such as density, pH and temperature are the main quality control parameters, which are usually measured once or twice a day to ensure a correct progression of the process and to avoid stuck and sluggish fermentations or contamination-related processes, which may lead to low quality wines [2]. If unexpected deviations occur, more exhaustive off-line laboratory analyses are needed, which involve delayed results that may not allow readjusting the process when it could still be solvable. Process Analytical Technologies (PAT) are based on the idea that quality of a product should be evaluated throughout the manufacturing process, by performing real-time measurements during processing instead of carrying out quality control measurements in the final product. PAT methodologies ensure that if a product operates under Normal Operation Conditions (NOC) it will probably meet the final quality requirements at the end of the process. PAT guidelines are founded on process understanding together with the fact that modern process analyzers can provide non-destructive measurements containing information related to biological, physical, and chemical attributes of the materials being processed [3]. Despite being developed for the pharmaceutical manufacturing, PAT have been gaining ground in the food and beverages industries [4]. In particular, when dealing with wine alcoholic fermentation monitoring and process control, the implementation of fast analytical tools, such as vibrational spectroscopy, has gained popularity over the last decades. Vibrational spectroscopy falls into the PAT guidelines as it allows getting real-time information of the process and taking corrective measures, if necessary, before obtaining the final product [5]. Among the different vibrational spectroscopy options, attenuated total reflectance mid-infrared spectroscopy (ATR-MIR) is a very valuable PAT tool for food and beverages analysis, as it is a fast and easy-to-use technique, which requires little or no sample pre-treatment [6].

To obtain the useful process information, the use of vibrational spectroscopy involves the acquisition of multivariate data and so it implies the application of multivariate statistical process control (MSPC) techniques. Among the different MSPC charts, the ones based on Principal Components Analysis (PCA) to monitor fed-batch processes are simple to represent and easy to interpret [7]. However, fed-batch processes naturally present several features, which make the modelling of NOC (Normal Operation Conditions) batches a

difficult task (e.g. time-varying dynamics and uneven batch length), and different approaches can be adopted depending on the type of process followed and the type of faults sought [8]. In batch processing, data can be represented as a three-way matrix of dimensions IxJxK (where I is the number of samples, J refer to number of variables and K is the number of time points of each batch). A decomposition into a two-way matrix of dimensions IJxK or IxJK can be applied to build multivariate PCA models. The basis of MSPC is very similar to the traditional univariate SPC methods, where the confidence limits are built based on data obtained only from NOC batches. To study the evolution of new batches, the statistical information from the PCA model is used. Particularly, Q residuals and Hotelling's T^2 values (calculated under normal distribution assumptions) are the most used statistical measures to detect irregular batches. Q residuals represent the squared perpendicular distance of a sample at a specific time point from the reduced space defined by the PCA model. They become greater when a batch deviates over time from NOC batches. Then, the irregular batch, when projected, lies outside the model, perpendicular to the NOC PCA space. In turn, Hotelling's T^2 values provide information of how far a batch is from the centre of the NOC reduced space. In this case, an abnormal batch would be positioned further away from the centre of the model as the deviation becomes more evident [9].

The combination of vibrational spectroscopy and MSPC tools to detect deviations during fed-batch processes in the food and beverages industries has already been considered, confirming the growing interest to integrate fast analytical tools and MSPC techniques into the process control line. Using FT-NIR, disturbances during the coffee roasting process were detected outside the limits of both T^2 and Q residual charts before the end of the process [10]. Similarly, faulty batches during the renneting process of milk were detected in the Q residuals chart [11]. Also, Q charts were used to detect off-specification coffee beans during storage in different packaging conditions using Raman spectroscopy [12]. However, information is very limited on the use of spectroscopic data and MSPC charts for on-line monitoring of fermentation processes in the agro-food sector to provide early indications of process deviations [13,14].

It has already been shown that ATR-MIR is suitable for real-time bioprocess monitoring [15] and particularly for monitoring industrial alcoholic fermentation processes [16,17]. In the winemaking industry, alcoholic fermentation monitoring using ATR-MIR has been widely studied and a review on the usefulness of this technique for process control can be found elsewhere [18]. Using ATR-MIR with PLSR, good prediction models were obtained

for alcohol, reducing sugars and titratable acidity during the alcoholic fermentation process in rice wine [19]. In another study, the prediction of glucose, fructose and glycerol during fermentation in red wine was achieved obtaining a r_{cv} (coefficient of correlation in cross-validation) above 0.99 [20]. The prediction of ethanol levels in several alcoholic beverages has also been reported, obtaining models with a root mean square error of prediction of 0.1% (w/w) [21]. Furthermore, it has been reported that ATR-MIR can detect deviations from NOC, including sluggish fermentations and microbiological spoilage [22,23], suggesting that this tool could be used for process control. From an oenological point of view, among the possible microbiological contaminations, those promoted by lactic bacteria are especially interesting because these lead to malolactic fermentation with the subsequent increase of the pH values (the diprotic malic acid is transformed into the monoprotic lactic acid, resulting in a deacidification of the wine). Therefore, if this process is not controlled, it gives rise to defective wines, in terms of its organoleptic profile. The most common preventive action to avoid this microbiological contamination is adding sulphites into the must at the beginning of alcoholic fermentation. However, health risks have been associated with sulphites so, the addition of sulphites only when a contamination is detected would be highly beneficial in health terms. In this case, the implementation of real-time monitoring in the cellar, together with MSPC charts, could be more efficient than performing off-line laboratory analyses, which provide delayed results and may not allow taking corrective measures when a deviation could still be solvable. Yet, the potential of MSPC charts in this field has not yet been fully investigated [24].

The aim of this study was to develop MSPC control charts as a tool for spoilage detection in the wine alcoholic fermentation process. This spoilage was promoted by inducing an additional malolactic fermentation (MLF) in some wine fermentations to evaluate the capability of the MSPC charts to detect this deviation from the normal process. The implementation of fast analytical tools such as ATR-MIR for process control would be particularly useful when dealing with microbiological contamination, as it could allow to shift from a preventive approach to a corrective one.

2. Materials and methods

2.1 Samples

The grape must employed to perform the small-scale fermentations (microvinifications) was obtained by the adequate dilution of a concentrated white grape must from Mostos Españoles S.A. (Ciudad Real, Spain). The diluted sugar (glucose and fructose)

concentrations were 200 ± 10 $g \cdot L^{-1}$, in order to reproduce natural variability in samples. In addition, yeast assimilable nitrogen was adjusted by supplementation with 0.30 $g \cdot L^{-1}$ of ENOVIT® (SPINDAL S.A.R.L Gretz Armainvilliers, France) and 0.30 $g \cdot L^{-1}$ of Actimaxbio* (Agrovin, Ciudad Real, Spain).

Microvinifications were conducted in 500 mL conical flasks by adding 350 mL of diluted must. Each flask was inoculated with the commercial dry yeast strain *Saccharomyces cerevisiae* "E491" (Vitilevure Albaflor, YSEO, Danstar Ferment A.G., Denmark), to reach a final concentration of $3 \cdot 10^6$ $CFU \cdot mL^{-1}$. In total, 25 NOC microvinifications were carried out in 4 different experiments throughout the year. To emulate the variability due to a real grape ripening process, each one of the four microvinification experiments used a must with a slightly different sugar concentration. Simultaneously to the NOC experiments, 8 additional microvinifications were intentionally contaminated with a freeze-dried blend of Lactic Acid Bacteria (*Lactobacillus plantarum* and *Oenococcus oeni*) in two different concentrations, 4 x $2.5 \cdot 10^6$ and 4 x $4 \cdot 10^6$ $CFU \cdot mL^{-1}$, to promote malolactic fermentation at different time points of the alcoholic fermentation. Samples were coded as MLF1 and MLF2, respectively. Rehydration of the microorganisms before co-inoculation was done following the suppliers' indications.

All the microvinifications were kept under constant temperature of 18 ºC until the end of the fermentations. Both alcoholic and malolactic fermentations were controlled by routine analysis twice a day until the end of both fermentations in order to ensure the normal progress of both processes (we considered that alcoholic fermentation ended when density was under 0.995 $g \cdot L^{-1}$ and malolactic fermentation ended when L-malic acid concentration was under 0.06 $g \cdot L^{-1}$). Alcoholic fermentations were controlled with density measurements with a portable densimeter (Densito2Go, Mettler Toledo, United States). Regarding to the malolactic fermentations, these were controlled by determining the L-malic acid concentration using a Y15 Analyser (Biosystems, Barcelona, Spain). pH was also continuously measured in both fermentations using a portable pH-meter with a 201 T portable electrode (7+ series portable pH-meter, XS Instruments, Italy). All the analyses were performed right after sample collection.

2.2 ATR-MIR analysis

After homogenization, 1,5 mL were collected at least once a day and centrifuged at 10000 rpm for 10 minutes, to avoid the scattering effect produced by the microorganisms present

in the sample. The pellet was discarded, and the supernatant was kept in 1.5 mL eppendorfs for further analysis. Infrared measurements were performed with a portable 4100 ExoScan FTIR instrument (Agilent, California, USA), equipped with an interchangeable spherical ATR sampling interface, consisting on a diamond crystal window. Spectra were collected using our previously optimized methodology [14] over the range 4000 to 650 cm^{-1} (resolution 8 cm^{-1}; 32 scans; triplicate per sample; air-background before sample). A drop of the sample was placed on top of the crystal and the spectrum was recorded immediately afterwards. Spectra were obtained using Microlab PC software (Agilent, California, USA) and data was saved as .spc files. The mean of the triplicates was used in subsequent data analysis.

2.3 Multivariate statistical process control

The spectral region selected to proceed with the study was from 1309 to 1082 cm^{-1}, which according to our previous studies is the region related to the malolactic fermentation [23]. The collected data consisted of a three-way matrix containing the absorbance values at different wavenumbers (J=62), for NOC and MLF samples (I=33) and at different sampling times (K) depending on the batch. Sampling times ranged from 0 to 210 hours, when the completion of both alcoholic and malolactic fermentations was achieved. Then, a time-wise unfolding of the three-way array was performed, resulting in a matrix with dimension IKxJ. After that, different pre-processing strategies where tested and optimized, including first and second derivatives, Savitzky-Golay smoothing and Standard Normal Variate (SNV). After spectral pre-processing, data were mean-centered.

First, a preliminary PCA model was built using only the data from the first experiment, in order to qualitatively visualize the main changes in the spectra and to detect trends in sample types (NOC and MLF). This model allowed to explore the variability of batches from the same experiment and the variability associated to the malolactic fermentation.

Next, three different strategies were applied with the unfolded, two-dimensional (IKxJ) matrix. In the first one, only NOC batches from a sole experiment were used to build a NOC PCA model (matrix of 340 rows – samples x time, and 62 columns - wavelengths). In the second strategy, NOC batches from the other experiments were added to build a new NOC PCA model including the variability among experiments. Thus, the 10 NOC samples from the previous model were used, but the matrix was augmented using 15 more NOC samples from the others three different experiments. The final NOC matrix consisted of 771 rows (NOC samples coming from the 4 experiments x time) and 62 columns

(wavelengths). In both approaches, samples from all the sampling times were used. Finally, in the third strategy, eight-hour models (interval PCA models) were developed used all the NOC data. A scheme of the procedure followed for all models is shown in Figure 1.

In the three strategies, MLF samples (272 rows – samples x time, x 62 columns - wavelengths) were projected in the different NOC PCA models by calculating their scores in the NOC reduced space and using the loadings obtained from each model. The capability of the PCA models to detect a deviation during the process using the defined reduced space and the statistical performance of the MLF samples were evaluated. All models were validated by applying the Kennard-Stone algorithm [25] using half of the NOC samples in the calibration set to ensure that the whole NOC variability is represented.

For each one of three models, T^2 and Q control charts were built. A 95% Hotelling's T^2 confidence limit was calculated using the NOC calibration samples, and then NOC samples from the validation set and all MLF samples were plotted in the Hotelling's T^2 control charts, representing T^2 values vs time. Similarly, a 95% confidence limit was calculated for the Q residuals using the NOC calibration samples, and then NOC samples from the validation set and MLF samples were projected in the Q control charts, representing in this case Q values vs time.

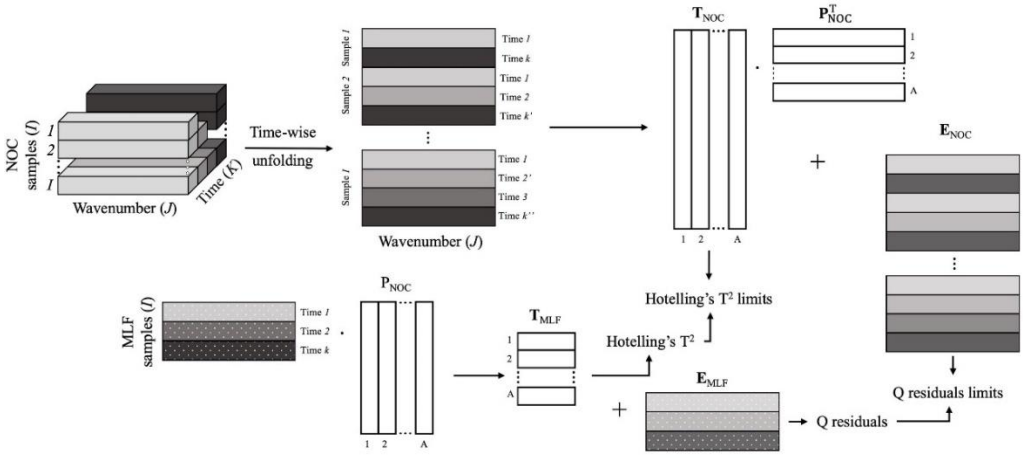


Figure 1. Scheme of the procedure applied to build the IKxJ PCA models for NOC samples and the projection of MLF samples.

All multivariate data analyses were performed using the PLS Toolbox v8.7 (Eigenvector Research Inc., Earglerock, USA) with MATLAB R2015b (The MathWorks, Natick, USA).

3. Results and discussion

3.1. Evolution of alcoholic and malolactic fermentations

As previously reported [26,27], the regions that show the greatest variability during alcoholic fermentation in the mid-infrared region are between 950 and 1500 cm^{-1}, due to sugars, acids, proteins and ethanol bonds absorption. In this region, the greatest variability associated to the biochemical transformation of sugars into ethanol and carbon dioxide is observed between 950 and 1100 cm^{-1}. However, when using the region from 950 to 1500 cm^{-1}, it was not possible to distinguish between NOC and MLF because spectral changes due to malolactic fermentation in this region were hidden by spectral changes corresponding to the main process (alcoholic fermentation) (data not shown). For this reason, to focus on the malolactic fermentation process, we used the region between 1309 to 1082 cm^{-1}, which maximised the differences between alcoholic and malolactic fermentations.

The optimized pre-processing used was first derivative (1st order polynomial) with Savitzky-Golay smoothing through 15 points, SNV and mean-center. Figure 2 (left) shows the raw mean spectra for NOC, MLF1 and MLF2 at the end of both alcoholic and malolactic fermentations, when the difference in the classes is maximum. Although very little separation emerges from the raw data, a decrease in absorbance can be seen between 1320 and 1170 cm^{-1}, which can be ascribed to a reduction in H-bonding as this region is associated with O-H bending vibrations. A difference in the slope can be appreciated in the region between 1150 and 1100 cm^{-1}, which is associated with stretching vibrations of C–C and C–O of carboxylic acids [28]. When applying our optimized pre-processing, the difference is noticeable and the differentiation between classes is detectable.

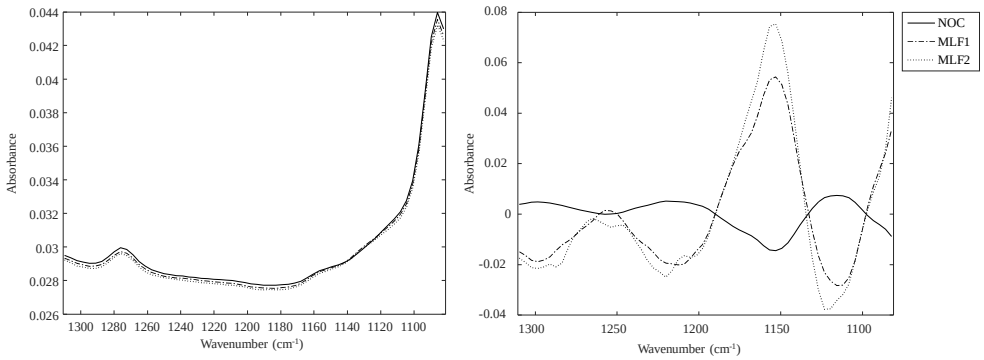


Figure 2. Raw mean spectra (left) and preprocessed mean spectra (right) for NOC, MLF1 and MLF2 at completion of alcoholic and malolactic fermentations

A preliminary PCA model was built only using samples from a single experiment (consisting on 10 NOC and 8 MLF batches), attempting to screen different behaviours between NOC and MLF samples. The number of principal components used for this purpose was optimized to 2 PCs, to well-define alcoholic fermentation and avoid overfitted models. As it can be observed in Figure 3, the first principal component (98.61% explained variance) follows the trend of alcoholic fermentation kinetics, as shown in our previous work [23]. From the second principal component (accounting for only a 1.04% of the total variability), a difference can be observed between NOC and MLF samples from hours 60-80 until the end of the process. It must be taken into account that the greatest variability in the spectra is due to the alcoholic fermentation, as it involves the transformation of sugars into ethanol from an initial concentration of 200 $g \cdot L^{-1}$. On the other hand, at the beginning of malolactic fermentation, malic acid concentration is only around 2 $g \cdot L^{-1}$ and the variability in the signal is much lower. Moreover, even though the second bioprocess does not interfere in the first bioprocess, sugars' and acids' bonds absorb in the same regions of the spectra, which makes it hard to detect MLF deviations. Figure 3 shows that differences between NOC and MLF start to be noticeable 80 hours after the beginning of the process, and a separation trend can be appreciated between 40 and 80 hours.

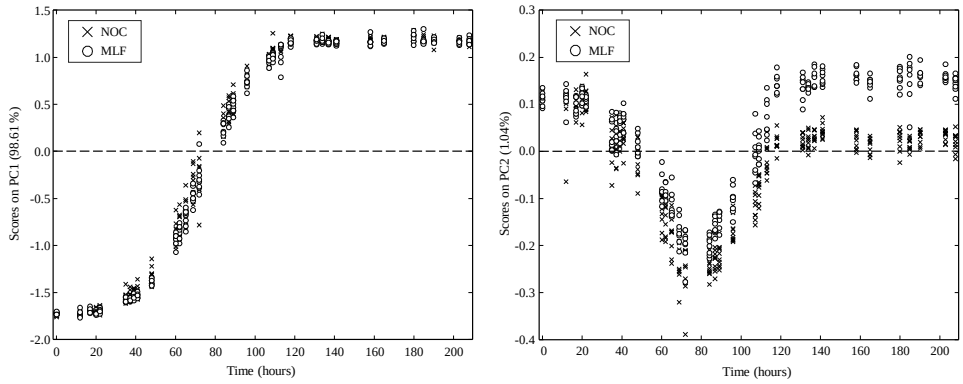


Figure 3. Scores for PC1 (left) and PC2 (right) for a singles batch fermentation.

3.2. MSPC charts for monitoring fermentations

3.2.1 Single experiment strategy

The MSPC charts used in this strategy are based on the Q and T^2 statistics. NOC samples used to build the PCA in Section 3.1 were also used to build a single experiment NOC PCA model (matrix 340x62). The scores of this model were used to calculate the T^2 95%

confidence limit, and the residual matrix of this model was used to calculate the Q-residuals 95% confidence limit. Q and T^2 vs time control charts are shown in Figure 4 for the 10 NOC and the 8 MLF batches (4 MLF1 and 4 MLF2). In both charts, NOC samples lie below the 95% confidence limit during the main course of the alcoholic fermentations. In contrast, MLF samples in the Q control chart show a significant difference from NOC samples from 80 hours onwards, as it was already observed in the PCA scores plot (Figure 3, section 3.1). On the other hand, all MLF samples lie below the confidence limits in the T^2 chart. Nevertheless, MLF samples exhibit a trend of higher T^2 values with respect to NOC samples from hour 110. These higher values could be explained because malolactic fermentation at this time is almost finished, and the distance to the centre of the model increases but not significantly, from a statistical point of view.

The fact that MLF samples are distinguishable from NOC samples in the Q chart but not in the T^2 chart is a reasonable result because the enormous variability in the spectra due to alcoholic fermentation especially between hours 70 and 90 when tumultuous fermentation takes place (in Figure 3, from hour 40 to 120, when sugar consumption is at its fastest rate) hampers the possibility to establish a confidence limit to differentiate the samples. It is important to remark that, despite the fact that a different trend between MLF and NOC samples can be seen after >100 hours, all samples fall under the confidence limit because of the mentioned variability among samples between hours 70 and 90 hours. Furthermore, malolactic fermentation evolution in the spectrum is jumbled with alcoholic fermentation, explaining the difficulty in finding the differentiation.

This methodology was validated by applying the Kennard-Stone algorithm. NOC samples were split using 50% of them to build the model (calibration set) and the remaining 50% for validation. Q values from the validation NOC samples were under the new Q residual confidence limit. Similarly, MLF samples showed a similar behaviour as in the model built with all NOC samples.

This time-wise unfolding approach is proposed as an alternative to IxJK batch-wise unfolding [8], were the exact same number of sampling times is required in order to project new suspected samples. We previously reported the use of a time-wise approach to detect sluggish alcoholic fermentations [14]. Here, any spectrum from a MLF batch can be projected onto model, and the model is able to determine if the sample is under or above the confidence limit, with no need to neither follow its complete alcoholic fermentation, nor to have the same exact amount of sampling points during the process. To confirm this idea,

MLF samples from a single time point (hour 119), which were above the Q confidence limit, were projected solely and, as samples are independent of the time, they were placed above the Q confident limit.

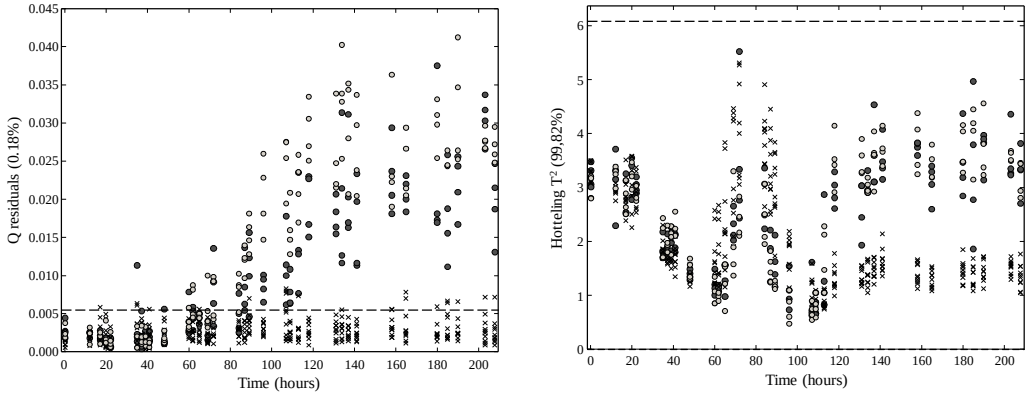


Figure 4. Q and T^2 charts for a single experiment. Symbols: (X) NOC, (●) MLF1, (○) MLF2.

3.2.2 Multiple experiment strategy

To confirm the hypothesis that it is not necessary to have the same sampling times and number of sampling points, three different experiments consisting of 15 additional NOC batches were added to the model. In every experiment, sampling was performed at similar, but not exactly, the same time points. This methodology perfectly agrees with the typical control timing in a winery, where the fermentation control is performed usually once or twice a day and not in each fermentation tank at the same time. In this case, as different NOC batches with different initial sugar concentrations (which implies different sugar consumption rates) were used (matrix dimensions of 771x62), the model was validated following the same procedure as in the single experiment model, assuring that NOC samples from all the experiments were included in both sets. As it can be observed in Figure 5, NOC validation samples are generally under the 95% confidence limit in both Q and T^2 control charts, assuring the validation of the model. In this model the trend observed in the Q control chart for the MLF samples is similar as in the previous PCA model when using a single experiment (Section 3.2), with a separation from hour 80 onwards. For the T^2 control chart, despite exhibiting the same trend observed in the model from Section 3.2, MLF samples are now statistically different from hour 120 until the end of the process. This may be explained because as more NOC samples are included in the NOC model, alcoholic fermentation variability is better described, and the model is able to detect the

differences among NOC samples and MLF, which can now be statistically differentiated after hour 120.

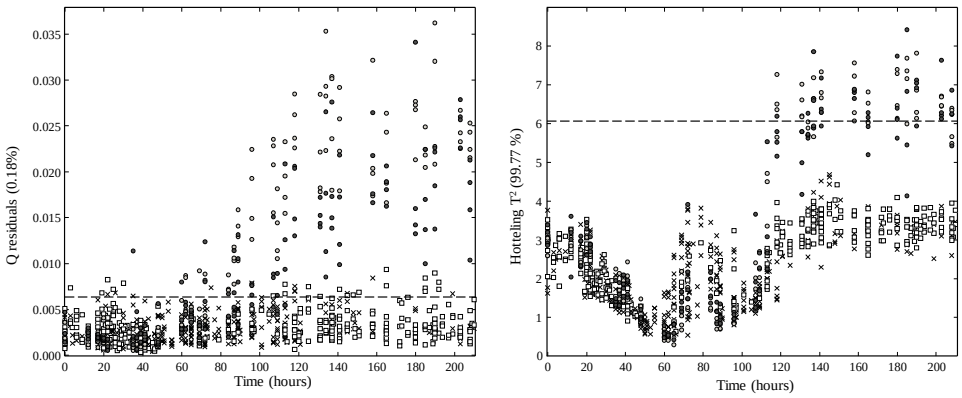


Figure 5. Q and T^2 charts for all the experiments. (X) Calibration NOC samples, (□) validation NOC samples, (●) MLF1, (○) MLF2.

As in the single experiment model, samples are independent of the sampling time and the full trajectory of the batch is not needed. Any single MLF spectrum can be projected into the model and this will be able to determine if the sample is statistically different in both Q and T^2 control charts. In other words, MLF samples will appear in the same position in the Q and T^2 charts as it happened when all the MLF samples were projected, regardless of the sampling times available.

3.2.3 Interval models' strategy

As it could be inferred from the T^2 control chart in Figure 5, during tumultuous fermentation (in Figure 3, from hour 40 to 120), higher T^2 values are observed. It would then be useful to build specific models for certain time intervals, as more accurate confidence limits could be stablished, especially for Hotelling's T^2. An 8-hour time interval was defined, because this period of time would be sufficient to take corrective measures in a winery and also, the changes in the matrix during this period would not affect the final product quality.

In this approach, all NOC and MLF matrix (771x61) was divided into 8-hour time-intervals, and PCA models were built for each interval until the separation of MLF samples. Matrixes of different sizes were used depending on the number of samples available at each time-interval. With the T^2 control charts, no statistical separation was achieved, showing that T^2 statistics does not represent a useful way to detect deviations. The model built with the time interval between 65 and 72 hours (Figure 6) shows a MLF separation trend. This

time interval model could be useful as an alert indicator before all MLF batches are completely separated, to foresee a spoilage at early stages of the malolactic fermentation. In this case, the Q values of MLF1 samples are slightly smaller than the ones of MLF2, as MLF2 samples have a higher concentration of LAB (higher L-malic acid consumption rate). The complete statistical separation in the Q chart was obtained in the model between 81 and 88 hours. Models were also built for subsequent time intervals, in order to assure that the separation is consistent until the end of the malolactic fermentation. Our results show that a malolactic fermentation detection threshold can be established as an indicator that a deviation is arising at the 65-72h interval, when 40-50% of this bioprocess has taken place. At this point, additional measures should be taken to readjust the principal process (alcoholic fermentation) and to avoid having a worse situation which could lead to wines with organoleptic defects or even worse, the loss of a whole vintage.

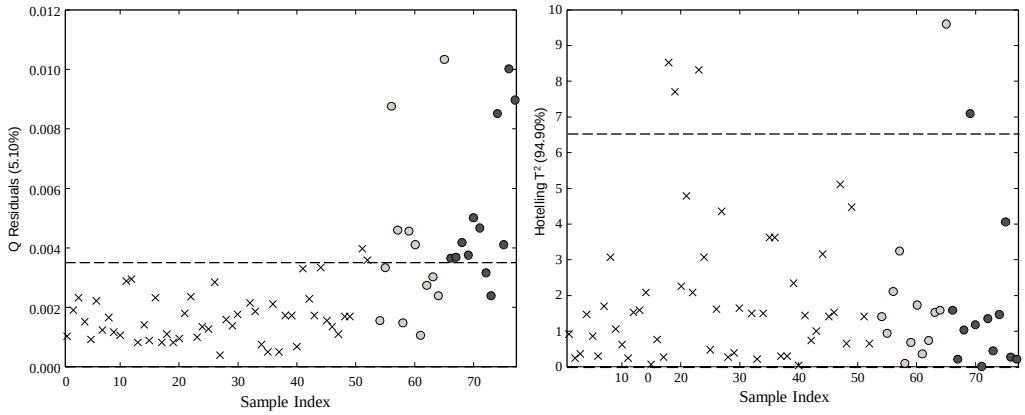


Figure 6. Time interval 65-72 hours. (X) NOC samples, (◯) MLF1, (●) MLF2.

4. Conclusions

In the best of our knowledge, this is the first time that Q residuals and Hotelling's T^2 control charts are used for the detection of an unwanted malolactic fermentation during alcoholic fermentation in wine based on ATR-MIR spectroscopic data. It was demonstrated that a specific signal pretreatment (e.g. batch alignment) is not required since typical year-to-year variability is considered in the global model. Also, using specific interval models improves the performance of the statistical detection of malolactic fermentation in the Q residual control chart. In conclusion, the different approaches here presented have the potential to be used in the oenological field, as an early detection of fermentation problems based on MSPC charts.

5. Acknowledgements

We thank the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (FEDER) (project AGL2015-70106-R, AEI-FEDER,UE) for the financial support given. Cavaglia, J. would also like to thank the Catalan Agency for Management of University and Research Grants (AGAUR) for the FI grant (FI_B100154).

6. References

[1] P. Ribéreau-Gayon, D. Dubourdieu, B. Donèche, A. Lonvaud, Handbook of Enology Volume 1 The Microbiology of Wine and Vinifications 2 nd Edition, 2nd Editio, 2006. doi:10.1002/0470010363.fmatter.

[2] L.F. Bisson, Stuck and sluggish fermentations, Am. J. Enol. Vitic. 50 (1999) 107–119.

[3] Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacuring, and Quality Assurance, FDA Off. Doc. (2004) 16.

[4] M. Jenzsch, C. Bell, S. Buziol, F. Kepert, H. Wegele, C. Hakemey, Trends in Process Analytical Technology: Present state in bioprocessing, in: B. Kiss, U. Gottschal, M. Pohlscheidt (Eds.), New Bioprocess. Strateg. Dev. Manuf. Recomb. Antibodies Proteins, 1st Editio, Springer International Publishing, Switzerland, 2018: pp. 211–252.

[5] F. Van Den Berg, C.B. Lyndgaard, K.M. Sørensen, S.B. Engelsen, Process Analytical Technology in the food industry, Trends Food Sci. Technol. 31 (2013) 27–35. doi:10.1016/j.tifs.2012.04.007.

[6] P. Roychoudhury, L.M. Harvey, B. McNeil, The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring, Anal. Chim. Acta. 571 (2006) 159–166. doi:10.1016/j.aca.2006.04.086.

[7] P. Nomikos, J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, Chemom. Intell. Lab. Syst. 30 (1995) 97–108. doi:10.1016/0169-7439(95)00043-7.

[8] J. Camacho, J. Picó, A. Ferrer, The best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?, Anal. Chim. Acta. 642 (2009) 59–68. doi:10.1016/j.aca.2009.02.001.

[9] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE J. 40 (1994) 1361–1375. doi:10.1002/aic.690400809.

[10] T.A. Catelani, J.R. Santos, R.N.M.J. Páscoa, L. Pezza, H.R. Pezza, J.A. Lopes, Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study, Talanta. 179 (2018) 292–299. doi:10.1016/j.talanta.2017.11.010.

[11] S. Grassi, L. Strani, E. Casiraghi, C. Alamprese, Control and Monitoring of Milk Renneting Using FT-NIR Spectroscopy as a Process Analytical Technology Tool, Foods. 8 (2019) 405. doi:10.3390/foods8090405.

[12] G.F. Abreu, F.M. Borém, L.F.C. Oliveira, M.R. Almeida, A.P.C. Alves, Raman spectroscopy: A new strategy for monitoring the quality of green coffee beans during storage, Food Chem. 287 (2019) 241–248. doi:10.1016/j.foodchem.2019.02.019.

[13] P. Jørgensen, J.G. Pedersen, E.P. Jensen, K.H. Esbensen, On-line batch fermentation process monitoring (NIR) - Introducing "biological process time," J. Chemom. 18 (2004) 81–91. doi:10.1002/cem.850.

[14] J. Cavaglia, B. Giussani, M. Mestres, M. Puxeu, O. Busto, J. Ferré, R. Boqué, Early detection of undesirable deviations in must fermentation using a portable FTIR-ATR instrument and multivariate analysis, J. Chemom. (2019) 1–11. doi:10.1002/cem.3162.

[15] J.J. Roberts, A. Power, J. Chapman, S. Chandra, D. Cozzolino, Vibrational Spectroscopy Methods for Agro-Food Product Analysis, in: Compr. Anal. Chem., 2018: pp. 51–68. doi:10.1016/bs.coac.2018.03.002.

[16] E.L. Veale, J. Irudayaraj, A. Demirci, An on-line approach to monitor ethanol fermentation using FTIR spectroscopy, Biotechnol. Prog. 23 (2007) 494–500. doi:10.1021/bp060306v.

[17] K.C.S. Rodrigues, J.L.S. Sonego, A. Bernardo, M.P.A. Ribeiro, A.J.G. Cruz, A.C. Badino, Real-Time Monitoring of Bioethanol Fermentation with Industrial Musts Using Mid-Infrared Spectroscopy, Ind. Eng. Chem. Res. 57 (2018) 10823–10831. doi:10.1021/acs.iecr.8b01181.

[18] R. Dambergs, M. Gishen, D. Cozzolino, A review of the state of the art, limitations, and perspectives of infrared spectroscopy for the analysis of wine grapes, must, and grapevine tissue, Appl. Spectrosc. Rev. 50 (2015) 261–278. doi:10.1080/05704928.2014.966380.

[19] D.Y. Kim, B.K. Cho, S.H. Lee, K. Kwon, E.S. Park, W.H. Lee, Application of Fourier transformmid infrared reflectance spectroscopy for monitoring Korean traditional rice wine "Makgeolli" fermentation, Sensor. Actuator. B Chem. 230 (2016) 753–760, doi:10.1016/j.snb.2016.02.076.

[20] Di Egidio V., Sinelli N., Giovanelli G., Moles A., Casiraghi E., NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, Eur. Food Res. Technol. 230 (2010) 947–955, doi:10.1007/s00217-010-1227-5.

[21] Debebe A., Redi-Abshiro M., Chandravanshi B.S., Non-destructive determination of ethanol levels in fermented alcoholic beverages using Fourier transform mid-infrared spectroscopy, Chem. Cent. J. 11 (2017) 1–8, doi:10.1186/s13065-017-0257-5.

[22] A. Urtubia, J. Ricardo Pérez-Correa, F. Pizarro, E. Agosin, Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems, Food Control. 19 (2008) 382–388. doi:10.1016/j.foodcont.2007.04.017.

[23] J. Cavaglia, D. Schorn-García, B. Giussani, J. Ferré, O. Busto, L. Aceña, M. Mestres, R. Boqué, ATR-MIR spectroscopy and multivariate analysis in alcoholic fermentation monitoring and lactic acid bacteria spoilage detection, Food Control. 109 (2020) 106947. doi:10.1016/j.foodcont.2019.106947.

[24] G. Hernández, R. León, A. Urtubia, Detection of abnormal processes of wine fermentation by support vector machines, Clust. Comput. 19 (2016) 1219–1225. doi:10.1007/s10586-016-0594-5.

[25] R.W. Kennard, L.A. Stone, Technometrics Computer Aided Design of Experiments, Technometric. 11 (1969) 137–148.

[26] D. Cozzolino, C. Curtin, The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments, Food Control. 26 (2012) 241–246. doi:10.1016/j.foodcont.2012.02.006.

[27] Z. Wu, E. Xu, J. Long, Y. Zhang, F. Wang, X. Xu, Z. Jin, A. Jiao, Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy, Food Control. 50 (2015) 405–412. doi:10.1016/j.foodcont.2014.09.028.

[28] Z. Richardson, D. Perez-Guaita, K. Kochan, B.R. Wood, Determining the age of spoiled milk from dried films using attenuated reflection fourier transform infrared (ATR FT-IR) spectroscopy, Appl. Spectrosc. 73 (2019) 1041–1050, doi:10.1177/0003702819842548.

Chapter 6

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Article 5

"Modelling and detection of lactic acid bacteria spoilage during wine alcoholic fermentation using ATR-MIR and MCR-ALS"

Manuscript to be submitted

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chemometric techniques are a valuable tool to extract information from spectroscopic data. In this chapter, spectroscopic data and MCR-ALS models are used to model both the alcoholic fermentation and the malolactic fermentation processes. We intended to address the last specific objective of the thesis, that is, whether MCR-ALS can be used to monitor and detect deviations during alcoholic fermentation, so that it could be applied as a PAT tool.

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) is a powerful chemometric technique that aims at resolving complex mixtures. One interesting advantage of MCR-ALS is that it is possible to obtain the pure spectra of the components involved in a chemical process, as well as the relative concentration (over time) of each component, only using spectroscopic data. Thus, to build informative models about the state of the process, there is no need of building an extensive database of reference chemical analyses, as it is required, for example to build robust PLSR models. In addition, the information from the MCR-ALS models can be used to build MSPC charts, developing a PAT method for the detection of lactic acid bacteria spoilage. Here we suggest the implementation of "inverse" MSPC charts, in which the model is built using the information from the specific deviation to be detected. In this way, using this type of MSPC we do not only detect a deviation, but we also can confirm that it is a deviation caused by the production of lactic acid through malolactic fermentation caused by a lactic acid bacteria spoilage.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Modelling and detection of lactic acid bacteria spoilage during wine alcoholic fermentation using ATR-MIR and MCR-ALS

Julieta Cavaglia[1], Silvia Mas Garcia[2,3], Jean-Michel Roger[2,3] Montserrat Mestres[1], Ricard Boqué[4]*

1. Universitat Rovira i Virgili. Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo s/n, Tarragona (43007) Spain
2. ITAP, Univ Montpellier, INRAE, Institut Agro, 34196 Montpellier, France
3. ChemHouse Research Group, 34196 Montpellier, France
4. Universitat Rovira i Virgili. Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo s/n, Tarragona, 43007, Spain

*Corresponding author: ricard.boque@urv.cat

Abstract

The aim of this study is to propose a new methodology based on the use of ATR-MIR spectra in combination with MCR-ALS to describe the evolution of the main components during alcoholic fermentation. Additionally, we propose a process control strategy base to detect changes between Normal Operation Conditions (NOC) fermentations and intentionally spoiled fermentations (MLF) with lactic acid bacteria at the beginning of alcoholic fermentation. In some wines, malolactic fermentation is avoided as it increases pH, which may have a detrimental effect on the final quality of the wine.

MCR-ALS models upon these data showed a good data fit (R^2= 99.95 and of lack of fit = 2.31%). It was possible to determine the spectral profiles from all relevant molecules, including the one related to bacterial spoilage (lactic acid). In addition, MSPC charts were built based on the concentration profiles obtained from MCR-ALS models, and using T^2 and Q statistics. Spoiled wines showed off-limit values for T^2 after 96 hours, making it possible to detect a lactic acid bacteria spoilage at early stages of alcoholic fermentation.

Alcoholic fermentation monitoring using ATR-MIR spectra and MCR-ALS analysis shows a great potential to rapidly control the state of a fermentation process, giving the possibility to detect the appearance of undesired molecules during the process and to apply corrective measures.

Key words: wine fermentation, bacterial spoilage, process control, Process Analytical Technologies (PAT), Multivariate Curve Resolution Alternating Least Squares (MCR-ALS).

1. Introduction

Wine alcoholic fermentation is a complex biological process that involves the transformation of grape must into wine by the action of yeasts. Although the main reaction is the conversion of reduced sugars into ethanol and CO_2, many other secondary products are obtained throughout the process, some of them of major enological importance such as organic acids, tannins and phenolic compounds, conferring the special characteristics to each wine [1].

The amount of ethanol and other secondary products in the final wine will depend on several factors, including the region were the vineyard is located, the grape variety, the yeast species used, the availability of nutrients and the fermentation conditions. Consequently, the high variability associated to this process makes alcoholic fermentation a difficult process to control [2]. In a winery, the daily measurement of important quality control parameters is required to ensure the correct evolution of the process. This usually includes density, pH and temperature. The sensory evaluation from an oenological expert may also play an important role. Nevertheless, additional chemical information during the process is needed in order to avoid unexpected deviations such bacterial spoilage. If the winery does not have its own analytical laboratory, wine samples during fermentation must be sent to an external analytical laboratory for additional off-line measurements of specific quality control parameters, which may mean getting delayed results and thus reducing the chances of taking correctives measures in time, especially when a deviation is suspected [3].

Vibrational spectroscopy has shown its suitability for the monitoring of alcoholic fermentation [4]. Using MIR spectroscopy during the fermentation process, good prediction values (R^2>0.99) were obtained for the most important quality control parameters in wine alcoholic fermentation monitoring, including glucose, fructose and the alcoholic degree [5–7]. Furthermore, the determination of other important parameters, such as total titratable acidity and total phenolic compounds have also been investigated, obtaining good models with low prediction errors [8,9].

Spectroscopic techniques, unlike traditional methods, are fast, nondestructive and usually require a minimum sample pretreatment [10]. These characteristics make vibrational spectroscopy of particular interest as a process analytical technology tool for bioprocess monitoring (e.g. alcoholic fermentation) [11]. Process analytical technologies (PAT) are a series of guidelines for the monitoring and quality control of products. They were first

introduced in 2004 by the American Food and Drugs Administration and aim at defining manufacturing processes through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials [12]. PAT tools have been gaining popularity for process and quality control of many food products, including wine [13].

One of the most important advantages of getting information in-time through the implementation of PAT methodologies, is that process deviations can be detected and, if possible, corrective measures before the end of the process can be taken [14]. In wine fermentation, bacterial spoilage may occur as different bacteria may be encountered in the winery environment [15]. For this reason, most wineries add sulphites into the fermentation tanks as a preventive measure, to avoid the growth of unwanted bacteria that may have a detrimental effect in the quality of the final wine [16]. Nevertheless, the use of sulphites is raising health concerns, and most wineries are trying to add as less sulphites as possible, or even to produce "sulphite-free" wines [17].

PLS is considered the most popular regression method for obtaining chemical information during alcoholic fermentation through spectroscopic data [18]. However, in order to obtain robust calibration models, it requires the analysis by traditional methods of key chemical parameters in a significant number of samples during the fermentation process, which is not always feasible [18]. In addition, food matrices consist of hundreds of organic compounds, making the analysis and interpretation of infrared spectra an extremely challenging task, due to the close similarity among the organic compounds, which results in overlapping bands [19].

Several studies have suggested the use of MCR-ALS applied to spectroscopic data as a monitoring tool during the fermentation processes. Grassi *et al.* have successfully applied MCR-ALS models to ATR-MIR data to describe the evolution of different fermentable sugars and ethanol during beer fermentation, obtaining a good data fit (percentage of explained variance >0.91%) [20]. In another study, milk fermentation was monitored applying MCR-ALS to FT-NIR data obtaining spectroscopic profiles for all the coagulation phases of milk and describing the main changes of milk during fermentation [21]. González-Sáiz *et al.* successfully described the relative concentration profiles and pure spectra during alcoholic fermentation of onion juice for sugars, ethanol and biomass using NIR data and MCR-ALS, reproducing 99.99% of the data and obtaining a low lack of fit (LOF = 0.09%) [22].

Several authors have considered the use of MCR-ALS in spectroscopic data as a PAT tool. Oliveira *et al.* implemented Multivariate Statistical Process Control (MSPC) charts based on MCR-ALS models to control distillation processes [23]. Using a 'Fixed size moving window strategy' and evolving MSPC charts, they obtained improved control charts with respect to the individual process observation models. In another study, Grassi *et al.* investigated the application of MCR-ALS to FT-NIR data from the milk renneting process. In their study, MSPC charts based on T^2 and Q statistics from PCA models built on MCR-ALS concentration profiles were able to distinguish coagulation problems in failure batches from the first minutes of the process [24].

Using MCR-ALS, the extraction of relevant information from MIR spectra on the evolution of the most relevant components during wine alcoholic fermentation would allow determining if the fermentation process is developing correctly. In fact, bacterial spoilage of wine is related to the appearance of certain molecules during fermentation. In the present study, we explored the application of MCR-ALS to monitor wine alcoholic fermentation through ATR-MIR data and to detect undesirable deviations during the process caused by the addition of spoilage bacteria. The main goals of this paper were two. Firstly, to describe, using MCR-ALS models, the different concentration evolutions of the main components in wine fermentation (sugars and ethanol) and the appearance of a third component (lactic acid) originating from a lactic acid bacteria spoilage through malolactic fermentation. Secondly, to evaluate if the MCR-ALS models obtained could be used as a process control tool to detect possible deviations during wine alcoholic fermentation.

2. Material and methods

2.1. Microvinifications

A batch of small-scale alcoholic fermentations (microvinifications) was performed using a concentrated white grape must (Mostos Españoles S.A, Ciudad Real, Spain). The must was defrosted and diluted to a final sugar (glucose and fructose) concentration of 200 ± 10 $g \cdot L^{-1}$. Microvinifications were performed in 500 mL conical flasks by adding 350 mL of diluted must. A total of 18 microvinifications were performed, 10 in Normal Operation Conditions (NOC), and 8 with an induced contamination of lactic acid bacteria (LAB). For all fermentations, yeast assimilable nitrogen was adjusted by supplementation with 0.30 $g \cdot L^{-1}$ of ENOVIT (SPINDAL, S.A.R.L. Gretz-Armainvilliers, France) and 0.30 $g \cdot L^{-1}$ of Actimaxbio* (Agrovin, Ciudad Real, Spain). All flasks were inoculated with the commercial

dry yeast strain *Saccharomyces cerevisiae* "E491" (Vitilevure Albaflor, YSEO, Danstar Ferment A.G., Denmark), reaching a final concentration of $3 \cdot 10^6$ CFU·mL^{-1}. To simulate the bacterial contaminations, the spoiled microvinifications were obtained by adding a freeze-dried blend of Lactic Acid Bacteria (*Lactobacillus plantarum* and *Oenococcus oeni*, Anchor Oenology, Montpellier, France) reaching two different concentrations: $2.5 \cdot 10^6$ and $4 \cdot 10^6$ CFU·mL^{-1} (for LAB1 and LAB2, respectively). Adding different concentrations of bacteria induces the deviations to start at different time points of the alcoholic fermentation.

2.2. Fermentation monitoring

All the microvinifications were kept at 18 ºC and density was measured by routine analysis at different time points until the end of alcoholic fermentation (density > 0.995 g·L^{-1}) to control the correct evolution of the process. After homogenization, 1.5 mL were collected at least once a day and centrifuged at 10000 rpm for 10 minutes, to avoid the scattering effect produced by the microorganisms present in the sample. The pellet was discarded, and the supernatant was kept in 2 mL eppendorfs for further analysis. Malolactic fermentation ended when L-malic acid concentration was under 0.06 g·L^{-1}. Density measurements were made with a portable densimeter (Densito2Go, Mettler Toledo, United States). L-malic acid was measured using a Y15 Analyser (Biosystems, Barcelona, Spain). All the analyses were performed right after sample collection.

2.3. ATR-MIR analysis

Infrared measurements were performed with a portable 4100 ExoScan FTIR spectrometer (Agilent, California, USA), equipped with an interchangeable spherical ATR sampling interface with a diamond crystal window. The spectroscopic range was from 4000 to 650 cm^{-1}, and spectra were recorded with a resolution of 8 cm^{-1} and 32 scans. An air-background was collected before each sample to avoid interferences due to the variation in room conditions. All samples were measured in triplicate. From each one of the 2 mL eppendorfs, a drop of the sample was placed on top of the crystal and the spectrum was recorded immediately afterwards. Spectra were collected using the Microlab PC software (Agilent, California, USA) and data was saved as .spc files. The mean of the triplicates was used in subsequent data analysis.

2.4. Data treatment

2.4.1. Data pretreatment

Data pretreatment is usually required to allow the extraction of chemical/physical information and attenuate undesirable signal contributions from the samples and/or the instrument. In the present work, a common preprocessing for spectral data was used: combination of Savitzky–Golay (SG) smoothing [25] with standard normal variate (SNV) scaling [26].

The smoothing parameters in SG (smoothing window of 15 points and first-order polynomial degree) were selected to keep the spectral features contained in the original spectra. Scattering effects were reduced by applying SNV scaling. Moreover, in order to minimize the influence of sugars in the data and because our objective was to focus on the absorptions from bonds of organic acids, the spectroscopic region between 1309 and 1082 cm^{-1} was selected for further analysis. This region falls in the fingerprinting region of the mid-infrared spectrum and is related to the absorption of several bonds that are characteristic of organic acids (C-O and C-C stretching, $-CH_2$ and $-CH_3$ bonds) [27,28].

2.4.2. Multivariate curve resolution-alternating least squares (MCR-ALS)

Every kth fermentation experiment monitored by ATR-MIR provides a data matrix, $\mathbf{D}_k$, where the rows are the spectra collected at different process times and the columns are the spectral wavelengths. The **D** matrix obeys the following bilinear model:

$$\mathbf{D} = \mathbf{C}\cdot\mathbf{S}^T + \mathbf{E} \qquad \textbf{Equation 1}$$

where **C** is the matrix of the kinetic profiles of the resolved compounds, $\mathbf{S}^T$ is the matrix of corresponding resolved pure spectra and the **E** matrix contains the experimental error or variance unexplained by the bilinear model.

In this case, in order to obtain more reliable results and the complete information about the different fermentations, all acquired ATR-MIR data should be treated together. For this purpose, all the $\mathbf{D}_k$ data matrices were arranged into a new augmented data matrix, setting one on top of each other and keeping the common wavelengths in the same column. The bilinear model in Equation 1 is now extended to the augmented data set as shown in Equation 2:

$$\mathbf{D}_{aug} = [\mathbf{D}_1;\mathbf{D}_2;\ldots;\mathbf{D}_k] = [\mathbf{C}_1;\mathbf{C}_2;\ldots;\mathbf{C}_n]\cdot\mathbf{S}^T + [\mathbf{E}_1;\mathbf{E}_2;\ldots;\mathbf{E}_k] = \mathbf{C}_{aug}\cdot\mathbf{S}^T + \mathbf{E}_{aug} \qquad \textbf{Equation 2}$$

where $\mathbf{C}_{aug}$ is a column-wise augmented matrix formed by the $\mathbf{C}_i$ submatrices that contain the resolved spectra at the different process times, and $\mathbf{S}^T$ is a single data matrix of pure spectra, assumed to be common and valid for all the different process times.

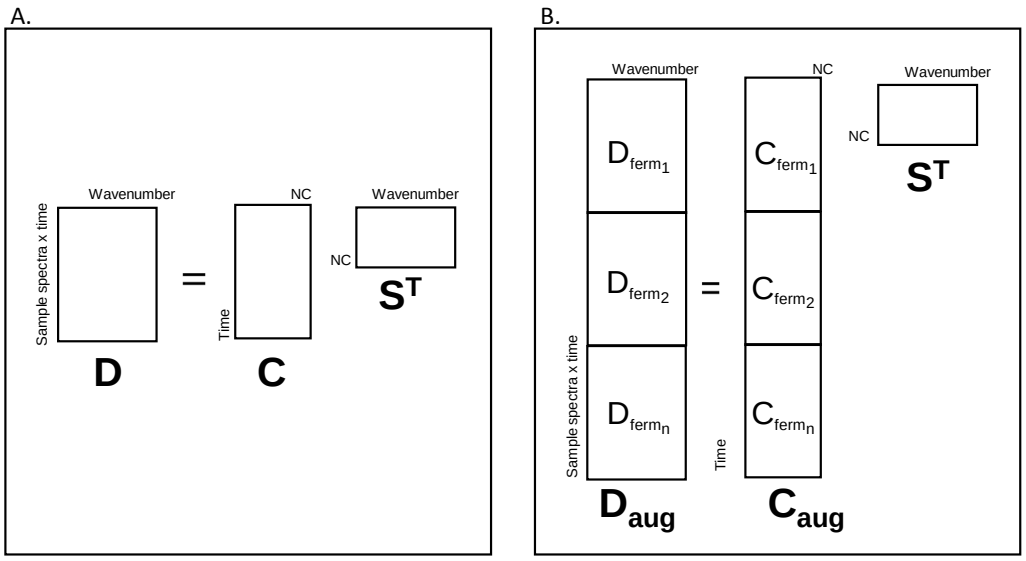


Figure 1. Bilinear model for the fermentation batches. **A.** Individual analysis for one single fermentation experiment. **B.** Simultaneous analysis including different fermentation batches to build the **C** augmented matrix ($\mathbf{C}_{aug}$). NC refers to the number of components.

Multivariate curve resolution-alternating least squares (MCR-ALS) aims at resolving the underlying bilinear model (see Equations 1 and 2) by using the sole information contained in the raw data set **D** [29,30].

MCR-ALS estimates iteratively the matrices $\mathbf{C}_{aug}$ and $\mathbf{S}^T$ by alternating least squares under the application of some constraints. In this study, the constraints applied to improve the final resolution and to minimize the MCR-ALS ambiguities were: non-negativity in the concentration profiles (concentration of the chemical compounds must be positive to have a physicochemical meaning), unimodality in the concentration profiles (presence of a single maximum per profile), normalization of pure spectra profiles and the constraint of correspondence among species to encode the information related to the presence/absence of some components in the different $\mathbf{C}_i$ submatrices.

The quality and reliability of the MCR-ALS models were evaluated by calculating two parameters that allow assessing the dissimilarity among the experimental data matrix (**D**) and the data modeled by MCR-ALS. These parameters are: the lack of fit (% LOF) and the explained variance (%r2), which are calculated using the following Equations:

$$\% \text{ LOF} = 100 \times \sqrt{\frac{\Sigma e_{ij}^2}{\Sigma d_{ij}^2}} \qquad \textbf{Equation 3}$$

$$\% \text{ r2} = 100 \times \left(1 - \frac{\Sigma e_{ij}^2}{\Sigma d_{ij}^2}\right) \qquad \textbf{Equation 4}$$

where d_{ij} is an element of the experimental data matrix **D** and e_{ij} is the related residual value obtained from the difference between the **D** matrix and the reproduced data ($\mathbf{C}\cdot\mathbf{S}^T$).

The number of components included in the MCR-ALS model is a compromise between model simplicity, maximum variance explained by the model, and model interpretability. MCR-ALS models were built using the MCR GUI (MCR UB, Barcelona) working under Matlab R2015 (The MathWorks, Natick, USA). More details about the MCR-ALS method are given in [31] and a GUI to use the algorithm is freely available at http://mcrals.info.

2.4.3. 'Inverse' MSPC charts

In process control involving multivariate data, MSPC charts are a valuable tool to monitor the effect of different variables at the same time. Due to the high amount of correlated data arising from spectroscopic data, MSPC methods are usually based on principal component analysis (PCA), where changes in the covariance structure of the process variables are detected. Conventional MSPC-PCA models identify process disturbances as soon as they occur using T^2 and Q statistical limits. Typically, a calibration model is developed using data collected from normal operating conditions (NOC) samples to define the design-space limits covering the NOC space, and deviations are detected when changes in the covariance matrix occur or abnormal signal components arise. The T^2 statistic calculates the distance from an observation to the center of the "in-control" set and determines whether a future observation has a systematic deviation in relation to the samples considered in statistical control. In turn, the Q statistic is defined as the square Euclidean distance perpendicular to an observation from the subspace defined by PCA, and it describes how well the PCA model predicts the recorded process variable [32].

In this study, the idea of 'inverse' MSPC charts is introduced. As explained in the introduction section, a fermentation deviation due to lactic acid bacteria spoilage is mainly due to the production of lactic acid. Thus, the modelling of lactic acid production cannot be performed only using NOC batches, as this molecule is not being produced during a normal alcoholic fermentation. In turn, the MCR-ALS model including the contamination samples must be considered, and all spectral profiles in the S^T matrix are used to obtain new C

matrices from spectroscopic data, which will be projected onto a PCA model only built with the C matrices of LAB fermentation batches. Figure 2 shows a scheme of the procedure.

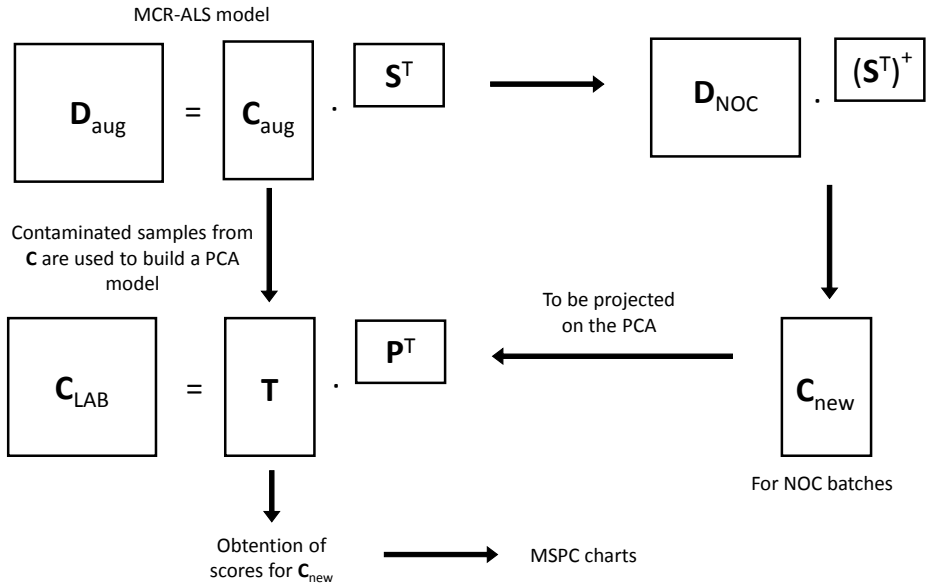


Figure 2. Scheme of the procedure followed to obtain the 'inverse' MSPC charts. The pseudoinverse of $\mathbf{S}^T$ ($(\mathbf{S}^T)^+$) is used to obtain a $\mathbf{C}_{new}$ matrix from spectroscopic data of control batches ($\mathbf{D}_{NOC}$).

3. Results and discussion

3.1 Fermentation monitoring

Figure 3 shows the mean evolution of the measured chemical parameters. Similar behaviors were found for each of the batches in their respective fermentation type. Density is an indirect measurement of the content of sugars in the must. Density curves show the typical sigmoidal form of sugar consumption. After 180 hours of fermentation, the consumption of sugars was completed for all microvinifications. In the present study, we used malic acid as an indirect measure of lactic acid production. The theoretical balance of malolactic fermentation states that 1 gram of malic acid consumed is transformed into 0.672 g of lactic acid and 0.328 g of CO_2. Thus, the approximate final concentration of lactic acid in the LAB fermentations, from the degradation of 1.6 g·L^{-1} of malic acid, is 1.075 g·L^{-1} [33].

All intentionally contaminated microvinifications produced lactic acid a few hours after the beginning of the alcoholic fermentation, confirming that those fermentations were deviated during the process. During the first 48 hours, changes in the concentration of malic acid were very slight. Between the third and fourth day of alcoholic fermentation, the conversion

rate of lactic acid increased. In LAB microvinifications, malolactic fermentation ended after 180 hours from the beginning of alcoholic fermentation. All NOC fermentations were maintained under control throughout the whole process, as less than 0.1 $g \cdot L^{-1}$ of malic acid was consumed.

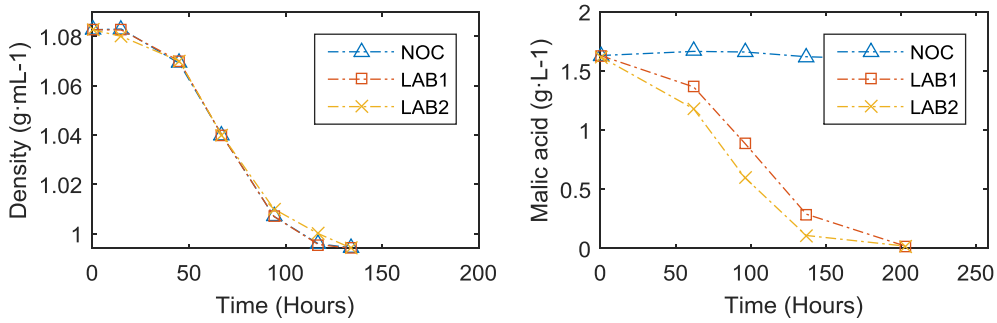


Figure 3. Evolution of the chemical parameters measured by reference analysis.

3.2 Multivariate Curve Resolution

The first MCR-ALS analysis was oriented to identify the specific contributions of both NOC and LAB fermentations. To this aim, two multisets (DaugNOC and DaugLAB) were built containing batches related to each particular type of fermentation. MCR-ALS was applied separately to each multiset structure using as constraints non-negativity and unimodality in the concentration profiles and normalization of the spectral profiles.

Table 1 lists the number of resolved components and the explained variance obtained from the MCR-ALS analyses of both multisets. Resolution of three contributions was necessary in both cases. No significant differences between resolved kinetic and spectral profiles of both fermentations (Figure not shown) were found. Therefore, it seems that there is a rank-deficiency phenomenon in LAB fermentations. Rank-deficiency implies that the number of components that can be modelled by MCR-ALS is lower than the actual number of chemical species involved in the reaction [34].

In this case, the rank deficiency in LAB fermentations could be due to the fact that malic acid and lactic acid have very similar spectra. To solve this problem, the analysis of a multiset containing all the batches (both NOC and LAB fermentations) was necessary. The different information present in the multiset structure enabled a significant improvement of the models, reducing the ambiguities associated and removing the rank deficiency.

Table 1. Number of resolved components, %LOF and variance explained by MCR-ALS analysis of $\mathbf{D}_{augNOC}$ and $\mathbf{D}_{augLAB}$ multiset structures.

Multiset	Resolved components	Lack of fit (%)	Explained variance (%)
$\mathbf{D}_{augNOC}$	3	1,97	99,96
$\mathbf{D}_{augLAB}$	3	1,27	99,98

So, MCR-ALS was applied to the multiset structure ($\mathbf{D}_{augNOCLAB}$) and four species were resolved, with a lack of fit (LOF %) equal to 2,3074 and 99,9488% of explained variance. The inclusion of a different number of species provided worse mathematical solutions or unreliable spectra for the concentration profiles. In this case, the constraint of correspondence of species was applied to encode the absence of lactic acid in NOC fermentations.

Figure 4 shows the $\mathbf{C}_{augNOCLAB}$ matrix for all fermentation batches. All batches showed similar concentration trends for all components, as expected. The curve that decreases during alcoholic fermentation (green curve) is related to sugars, which are consumed during alcoholic fermentation. Alcohol is produced from the consumption of sugars and is represented by the curve that increases (red line) a few hours after the beginning of the fermentation. Lactic acid was also simple to assign, given that only LAB fermentations showed the presence of this component (purple line). The curve showing a peak in the middle of the fermentation process (blue line) was more difficult to elucidate. It could be attributed to the presence of salts of tartaric acid, which precipitate by the end of alcoholic fermentation, or intermediate organic species that are part of the yeast metabolism.

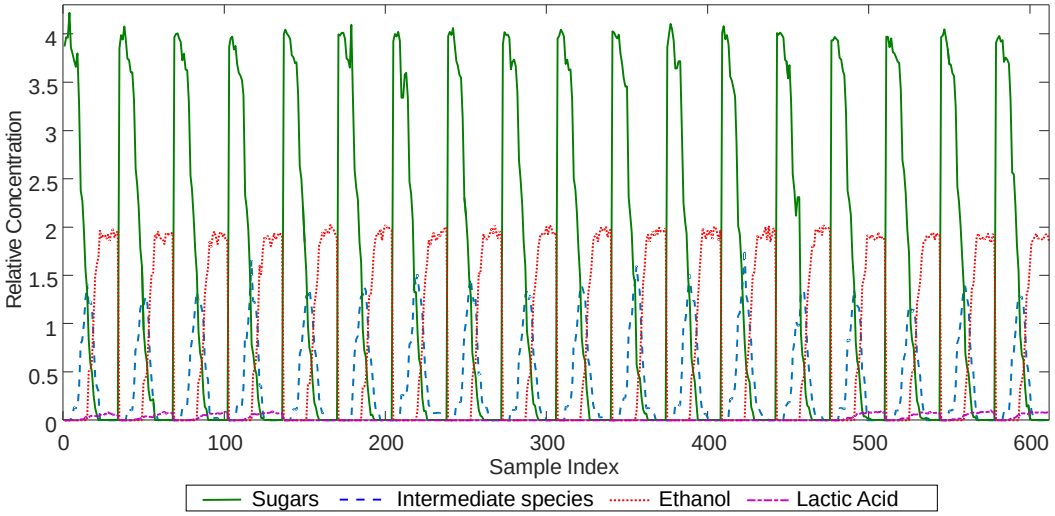


Figure 4. Relative concentrations for all fermentation batches. The reader is referred to the online version of the paper for legend colors.

In terms of proportion, the ratios obtained by the MCR-ALS are accurate, as the concentration of sugars goes from 200 to <0.05 $g \cdot L^{-1}$, while the concentration of lactic acid goes from zero to approximately 1.075 $g \cdot L^{-1}$, as calculated from theoretical conversion from malic acid. Our results show that MCR-ALS is able to find the relations between major components in wine (ethanol and sugars) and minor components such as lactic acid.

Figure 5 shows the pure signals ($\mathbf{S}^T$ matrix) obtained for the four components in the MCR-ALS model considering all batches (NOC and LAB microvinifications). Spectral profiles are appropriate as they show absorbance levels in the expected regions for organic compounds. Glucose and fructose were considered as a single species (sugars), due to the high amount of overlapping of the absorption bands in the MIR region for these molecules. Between 1089 and 1126 cm^{-1} the spectra are extremely overlapped. In this region C-C and C-H stretching vibrations are found, which are very common in organic molecules. A peak for lactic acid is observed at 1150 cm^{-1}, which can be ascribed to C-O stretching from carboxylic acids [35].

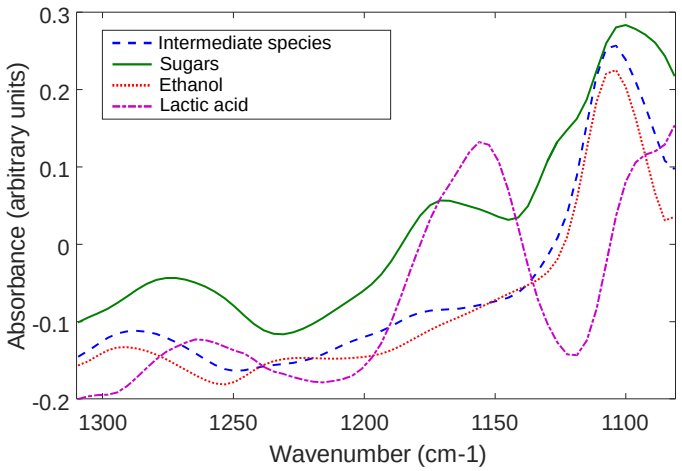


Figure 5. Pure spectral profiles for the four components in the resulting MCR-ALS model. The reader is referred to the online version of the paper for legend colors.

3.3 Inverse' MSPC charts

In this article, we propose the development of MSPC charts based on the contaminated samples, rather than on NOC samples. When a sample is projected into the model, if it falls under the "control" limits of the charts, it would mean exactly the opposite from a traditional MSPC chart. Instead of being under control, it would mean that lactic acid is

being produced and, therefore, the fermentation should be corrected to return to the normal conditions.

Following the scheme in Figure 2, a PCA model was built using the $\mathbf{C}_{augNOCLAB}$ matrix, containing all fermentation batches. The $\mathbf{S}^T$ matrix was used to create a new **C** matrix ($\mathbf{C}_{new}$) from the original NOC spectra. Then, the $\mathbf{C}_{new}$ matrix was projected onto the PCA model and the Hotelling T^2 values for these samples were included in the MSPC chart.

Four principal components (PCs) were used to build the PCA model, which explained 100% of the variability in the data, as the projection was made in a four-dimensional space correspoding to the four components resolved by MCR-ALS. Figure 6 shows the loadings of each PC in this model. In PC1, sugars account for most of the variation, while ethanol is the dominating compound in PC2. In PC3, the intermediate species and ethanol are the most important. Finally, the information on lactic acid is found in PC4.

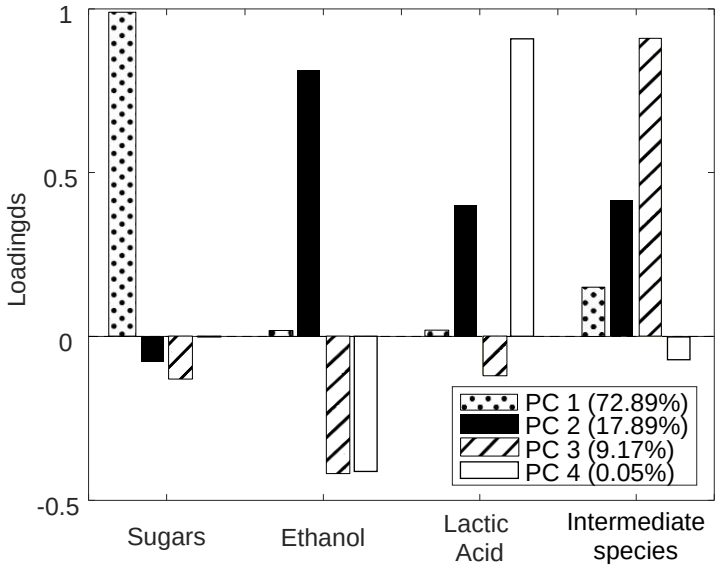


Figure 6. Loadings for the 4 PCs in the PCA model obtained from the **C** matrix of the MCR-ALS model.

The 'inverse' MSPC chart based on the Hotelling T^2 statistic is shown in Figure 7. In this MSPC chart, all samples under the control limits (round shapes) belong to LAB fermentations, were lactic acid is produced during alcoholic fermentation. Between 0 and 60 hours, all NOC samples are "under control". This could be explained because during the first 65 hours of alcoholic fermentation, the concentration curve for lactic acid is 0 for both LAC and NOC samples, as the production of lactic acid in this time interval is below the

limit of detection of the ATR-MIR instrument. Atypical T^2 values for NOC samples start to appear after 60 hours, as no lactic acid is being produced. All NOC samples are completely out of the limit after 96 hours. In our MSPC chart, four LAB samples were out-of-limit after 110 hours, and another LAB sample was off the limit at 190 hours.

Oliveira *et al.* used local PCA models and were able to build Fixed Size Moving Window MSPC charts and evolving MSPC charts to detect faulty batches during the distillation process using NIR data [23]. In our approach, faulty batches are detected using a global PCA model, with no need to build independent PCA models for different times.

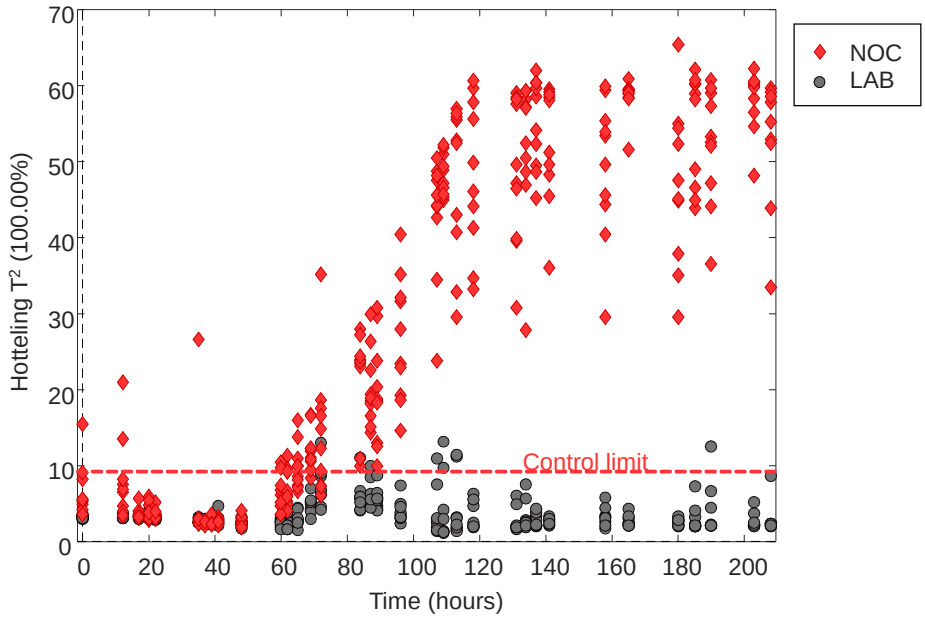


Figure 7. MSPC chart based on Hotelling T^2, showing the distribution of NOC samples (diamond shape) and LAB samples (round shape) throughout the alcoholic fermentation.

4. Conclusions

Our results suggest that ATR-MIR data together with MCR-ALS models and MSPC charts could be used for the detection of lactic acid production during alcoholic fermentation. The use of MCR-ALS with ATR-MIR spectra from wine alcoholic fermentation provides the possibility to model the pure kinetic and spectra profiles of the main compounds involved in wine alcoholic and malolactic fermentations. This methodology comes as an improvement of the traditional MSPC charts as. this way, when combined with the use of traditional MSPC charts for bacterial spoilage detection (as suggested in Cavaglia *et al.* [36]), if a fermentation batch is out of control in a traditional

MSPC batch, but in control in this new 'inverse' MSPC chart, we could not only conclude that the sample is deviated, as previously reported, but we could also state that the fermentation is deviated because of the production of lactic acid, as shown in the relative concentration profiles of MCR-ALS. Using 'inverse' MSPC charts it has been possible to detect spoiled wines due to lactic acid bacteria before the end of alcoholic fermentation. Also, as no prior information of the process is required, spectroscopic data could be obtained as many times as possible in a real winery environment with no need of reference analysis nor previously validated calibration models, giving the possibility to gather a lot of data at a small cost.

5. Aknowledgements

Authors would like to thank the scientific advice received from Anna de Juan (Universitat de Barcelona, Spain). Julieta Cavaglia would like to thank to the ChemHouse Research Group (Montpellier, France) for accepting her as a visiting research fellow. Financial support is acknowledged for the *Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya*, the European Union (UE) and the European Social Fund (ESF) (predoctoral fellowship number 2018 FI_B 00844), and the Spanish Ministry of Science and Technology (Project AGL2015-70106-R, AEI/FEDER, UE).

6. Bibliography

[1] P. Ribéreau-Gayon, D. Dubourdieu, B. Donèche, A. Lonvaud, Handbook of Enology, Vol. 1: The Microbiology of Wine and Vinifications, 2nd Editio, John Wiley & Sons, 2006. https://doi.org/10.1002/0470010363.fmatter.

[2] J.M. Sablayrolles, Control of alcoholic fermentation in winemaking: Current situation and prospect, Food Res. Int. 42 (2009) 418–424. https://doi.org/10.1016/j.foodres.2008.12.016.

[3] L.F. Bisson, C.E. Butzke, Diagnosis and rectification of stuck and sluggish fermentations, Am. J. Enol. Vitic. (2000).

[4] C.A.T. dos Santos, R.N.M.J. Páscoa, J.A. Lopes, A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle, Trends Anal. Chem. 88 (2017) 100–118. https://doi.org/10.1016/j.trac.2016.12.012.

[5] A. Urtubia, J.R. Pérez-Correa, M. Meurens, E. Agosin, Monitoring large scale wine fermentations with infrared spectroscopy, Talanta. 64 (2004) 778–784. https://doi.org/10.1016/j.talanta.2004.04.005.

[6] D. Cozzolino, W. Cynkar, N. Shah, P. Smith, Feasibility study on the use of attenuated total reflectance mid-infrared for analysis of compositional parameters in wine, Food Res. Int. 44 (2011) 181–186. https://doi.org/10.1016/j.foodres.2010.10.043.

[7] P. Fayolle, D. Picque, B. Perret, E. Latrille, G. Corrieu, Determination of Major Compounds of Alcoholic Fermentation by Middle-Infrared Spectroscopy: Study of Temperature Effects and Calibration Methods, Appl. Spectrosc. 50 (1996) 1325–1330. https://doi.org/10.1366/0003702963904872.

[8] S. Fragoso, L. Aceña, J. Guasch, M. Mestres, O. Busto, Quantification of phenolic compounds during red winemaking using FT-MIR spectroscopy and PLS-regression, J. Agric. Food Chem. 59 (2011) 10795–10802. https://doi.org/10.1021/jf201973e.

[9] D. Picque, D. Lefier, R. Grappin, G. Corrieu, Monitoring of fermentation by infrared spectrometry: Alcoholic and lactic fermentations, Anal. Chim. Acta. 279 (1993) 67–72. https://doi.org/10.1016/B978-0-444-81640-5.50010-0.

[10] J.J. Roberts, A. Power, J. Chapman, S. Chandra, D. Cozzolino, Vibrational Spectroscopy Methods for Agro-Food Product Analysis, in: Compr. Anal. Chem., 2018: pp. 51–68. https://doi.org/10.1016/bs.coac.2018.03.002.

[11] D. Landgrebe, C. Haake, T. Höpfner, S. Beutel, B. Hitzmann, T. Scheper, M. Rhiel, K.F. Reardon, On-line infrared spectroscopy for bioprocess monitoring, Appl. Microbiol. Biotechnol. 88 (2010) 11–22. https://doi.org/10.1007/s00253-010-2743-8.

[12] Food and Drug Administration, Guidance for Industry. PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, (2004). http://www.fda.gov/cvm/guidance/published.html (accessed September 2, 2020).

[13] C. O'Donnell, C. Fagan, P.J. Cullen, Process Analytial Technology for the Food Industry, 2009. https://doi.org/10.1016/S1571-5078(08)00422-4.

[14] M. Jenzsch, C. Bell, S. Buziol, F. Kepert, H. Wegele, C. Hakemey, Trends in Process Analytical Technology: Present state in bioprocessing, in: B. Kiss, U. Gottschal, M. Pohlscheidt (Eds.), New Bioprocess. Strateg. Dev. Manuf. Recomb. Antibodies Proteins, 1st Editio, Springer International Publishing, Switzerland, 2018: pp. 211–252. http://www.springer.com/series/10.

[15] A. Barata, M. Malfeito-Ferreira, V. Loureiro, The microbial ecology of wine grape berries, Int. J. Food Microbiol. 153 (2012) 243–259. https://doi.org/10.1016/j.ijfoodmicro.2011.11.025.

[16] C.S. Ough, Determination of sulfur dioxide in grapes and wines., in: J. Assoc. Off. Anal. Chem., 1986: pp. 5–7. https://doi.org/10.1093/jaoac/69.1.5.

[17] C. Ubeda, R. Hornedo-Ortega, A.B. Cerezo, M.C. Garcia-Parrilla, A.M. Troncoso, Chemical hazards in grapes and wine, climate change and challenges to face, Food Chem. 314 (2020) 126222. https://doi.org/10.1016/j.foodchem.2020.126222.

[18] S. Wold, M. Sjostrom, L. Eriksson, S. Sweden°°, PLS-regression: a basic tool of chemometrics, 2001. www.elsevier.comrlocaterchemometrics (accessed April 11, 2019).

[19] A.P. Craig, A.S. Franca, J. Irudayaraj, Vibrational spectroscopy for food quality and safety screening, in: High Throughput Screen. Food Saf. Assess. Biosens. Technol. Hyperspectral Imaging Pract. Appl., 1st Editio, Elsevier Ltd, 2015: pp. 165–194. https://doi.org/10.1016/B978-0-85709-801-6.00007-1.

[20] S. Grassi, J.M. Amigo, C.B. Lyndgaard, R. Foschino, E. Casiraghi, Assessment of the sugars and ethanol development in beer fermentation with FT-IR and multivariate curve resolution models, Food Res. Int. 62 (2014) 602–608. https://doi.org/10.1016/j.foodres.2014.03.058.

[21] S. Grassi, C. Alamprese, V. Bono, E. Casiraghi, J.M. Amigo, Modelling Milk Lactic Acid Fermentation Using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), Food Bioprocess Technol. 7 (2014) 1819–1829. https://doi.org/10.1007/s11947-013-1189-2.

[22] J.M. González-Sáiz, I. Esteban-Díez, S. Rodríguez-Tecedor, C. Pizarro, Valorization of onion waste and by-products: MCR-ALS applied to reveal the compositional profiles of alcoholic fermentations of onion juice monitored by near-infrared spectroscopy, Biotechnol. Bioeng. 101 (2008) 776–787. https://doi.org/10.1002/bit.21939.

[23] R.R. de Oliveira, R.H.P. Pedroza, A.O. Sousa, K.M.G. Lima, A. de Juan, Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy, Anal. Chim. Acta. 985 (2017) 41–53. https://doi.org/10.1016/j.aca.2017.07.038.

[24] S. Grassi, L. Strani, E. Casiraghi, C. Alamprese, Control and monitoring of milk renneting using FT-NIR spectroscopy as a process analytical technology tool, Foods. 8 (2019). https://doi.org/10.3390/foods8090405.

[25] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Anal. Chem. 36 (1964) 1627–1639. https://doi.org/10.1021/ac60214a047.

[26] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777. https://doi.org/10.1366/0003702894202201.

[27] D. Picque, P. Lieben, P. Chrétien, J. Béguin, L. Guérin, Assessment of maturity of loire valley wine grapes by mid-infrared spectroscopy, J. Int. Des Sci. La Vigne Du Vin. (2010). https://doi.org/10.20870/oeno-one.2010.44.4.1477.

[28] I. Vigentini, S. Grassi, N. Sinelli, V. Di Egidio, C. Picozzi, R. Foschino, E. Casiraghi, Near and Mid Infrared Spectroscopy to detect malolactic biotransformation of *Oenococcus oeni* in a wine-model., J. Agric. Sci. Technol. 4 (2014) 475–786.

[29] A. de Juan, S.C. Rutan, R. Tauler, MCR Capítols, in: R. Brown, B. Tauler, Walczak (Eds.), Compr. Chemom., Elsevier, Amsterdam, 2009: pp. 325–344. https://doi.org/10.1016/B978-044452701-1.00050-8.

[30] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. https://doi.org/10.1016/0169-7439(95)00047-X.

[31] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12. https://doi.org/10.1016/j.chemolab.2014.10.003.

[32] Q. Chen, U. Kruger, M. Meronk, A.Y.T. Leung, Synthesis of T^2 and Q statistics for process monitoring, Control Eng. Pract. 12 (2004) 745–755. https://doi.org/10.1016/j.conengprac.2003.08.004.

[33] A. Larrea Redondo, Enología básica, in: Enol. Básica, AEDOS EDITORIAL S.A., Barcelona, 1982.

[34] M. Blanco, M. Castillo, A. Peinado, R. Beneyto, Application of Multivariate Curve Resolution to Chemical Process Control of an Esterification Reaction Monitored by Near-Infrared Spectroscopy, n.d.

[35] J. Chapman, S. Gangadoo, V.K. Truong, D. Cozzolino, Spectroscopic approaches for rapid beer and wine analysis, Curr. Opin. Food Sci. 28 (2019) 67–73. https://doi.org/10.1016/j.cofs.2019.09.001.

[36] J. Cavaglia, D. Schorn-García, B. Giussani, J. Ferré, O. Busto, L. Aceña, M. Mestres, R. Boqué, Monitoring wine fermentation deviations using an ATR-MIR spectrometer and MSPC charts, Chemom. Intell. Lab. Syst. 201 (2020) 104011. https://doi.org/10.1016/j.chemolab.2020.104011.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chapter 7

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

General discussion

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

In this thesis, we investigated the potential use of a portable ATR-MIR spectrometer, in combination with chemometric techniques, for the detection of undesirable deviations during wine alcoholic fermentation

7.1. Instrumentation set-up

At instrumental level, in our research the optimal spectral resolution was 8 cm^{-1}, as we found that increasing the resolution did not add any relevant information to the spectra. Other authors have also reported that this resolution is sufficient to efficiently study the composition of wines and other beverages[1,2]. As for the number of scans, depending on the instrument and the goal of the study, different number of scans can be necessary to improve the outcomes of the multivariate data analyses. Even though increasing the number of scans improves the resolution of the spectra, the choice of the number of scans is a trade of the time required per measurement off for the resolution. For our fermentations, we found that 32 scans were enough to build robust models. In contrast, Kölhed and Karlberg reported that 64 scans were more suitable for the identification and quantification of natural sugars (glucose, fructose and sucrose) in fruit juices using an on-line capillary electrophoresis system coupled to Fourier Transform infrared spectroscopy[3].

From a qualitative point of view, it was possible to monitor the evolution of the whole alcoholic fermentation process by following the changes of the ATR-MIR spectra samples throughout the fermentation, that is, including all time points (**Articles 1, 2 and 3**). The main changes were observed in the fingerprinting region between 950 and 1500 cm^{-1}. In this region, organic acids and sugars show important absorption peaks. Specifically, deformations of $–CH_2$, C–C–H and H–C–O can be found between 1500 and 1200 cm^{-1} whereas C–C and C–O stretching vibrations are found between 1200 and 950 cm^{-1}. In addition, a broad band between 3000 and 3500 cm^{-1} could be ascribed to water and ethanol O-H stretching vibrations, which also show variability during wine alcoholic fermentation due to ethanol content increase. These results are in agreement with the literature, both in ATR and transmission IR modes. Di Egidio *et al.* found that the main MIR regions showing variation during red wine alcoholic fermentation were 3700 to 3000, 1700 to 1400 and 1200 to 1000 cm^{-1}, which correspond to the absorption bands of O–H, C=O and C–C/C–O stretching vibrations[4]. Using an ATR-MIR, Schalk *et al.* also reported the spectral regions from 1200 to 950 cm^{-1} as the main regions where the consumption of glucose and the simultaneous formation of ethanol are predominantly reflected during aerobic yeast fermentation in a culture medium. Similar trends could be observed in the range from 3400

to 2600 cm^{-1}, but with lower signal intensities[5]. In **Article 3**, the region from 850 to 649 cm^{-1} was excluded, as it did not contain useful information and it added noise into the models. Additionally, a variable selection process was applied for the detection of lactic acid bacteria (LAB) spoilage. Some authors also reported the need of excluding some spectroscopic regions that are not related to any chemical information and would add unwanted variation to the data, such as the regions from 830 cm^{-1} downwards and 3627 cm^{-1} onwards[6-8]. In addition, some authors have found that using water as the background spectra (prior to sample spectra acquisition) provides better results for the determination of some quality parameters such as ethanol or phenolic compounds[9,10]. In our case, however, using water as background did not improve the models, and instead an air background was taken each time. Other authors have also used air as background to build chemometric models on beverage samples using ATR-MIR and obtained satisfactory results[11,12]. Therefore, the choice of the type of background should be studied in each case depending on the goal of the study and the instrument characteristics.

7.2. Spectral preprocessing

One important aspect that must be addressed before any MVDA analysis is spectral preprocessing. From our results in **Article 1** and **Article 2**, we found that the information arising from the spectra could improve depending on the preprocessing strategy applied. For the prediction of chemical parameters, we applied a minimum preprocessing: Savintzky-Golay Smoothing, Standard Normal Variate (SNV) normalization and mean centering, in order to maintain the structure of original data as much as possible and facilitate chemical interpretation, which is one of the advantages of using MIR spectra. In the literature, there is not a consensus on the use of specific preprocessing methods for MIR spectra in the analysis of foodstuffs and beverages. First derivative is one of the most common preprocessing methods, followed by SNV, second derivative and the combination of first derivative and Multiplicative Scatter Correction (MSC)[13]. Some authors, however, have proposed other alternatives. Vector normalization and no-preprocessing of spectral data was suggested to obtain suitable Partial Least Squares Regression (PLSR) models with low prediction errors (RMSECV) in the analysis of free and copigmented anthocyanins, polymeric pigment fractions and actual red color in wine samples[14]. Cozzolino and Curtin used raw spectra with no mathematical preprocessing to predict several chemical parameters in wine. They obtained low standard errors of prediction (SEP) for ethanol (0.11%), pH (0.10), titratable acidity (0.53 $g \cdot L^{-1}$) and glucose + fructose (1.35 $g \cdot L^{-1}$)[15]. With our data, we found that the absence of spectral preprocessing did not yield satisfactory

results, and the performance of PLSR models, in terms of RMSECV and RMSEP, improved substantially after preprocessing. One of the reasons for trying different preprocessing methods is that the validation and performance evaluation of PLSR models can vary greatly from one study to another. Some authors use cross-validation, while others use external datasets to validate the models.

7.3. Predictive analysis

In **Article 1**, a PLSR model was built to predict the concentration of sugars along the fermentation process, from the spectra of the fermenting samples. All data were used for model calibration, and the cross-validation method was used for validation. The root mean squared error of calibration (RMSEC) obtained was 10.6 $g \cdot L^{-1}$ and the RMESCV was 10.9 $g \cdot L^{-1}$. Although traditional methods such as the enzymatic determination of sugars show lower error values, the simplicity and rapidity in obtaining results that offers an ATR-MIR instrument when using suitable chemometric tools, is an interesting alternative to provide valuable information on the status of the fermentation process. In **Article 2**, pH and density were measured throughout the fermentation process to investigate the prediction capability of the ATR-MIR spectrometer for these parameters. It was possible to predict density and pH with acceptable prediction errors using ATR-MIR spectra and PLSR. A variable selection step was introduced by applying the Selectivity Ratio algorithm. For the prediction of density, the spectroscopic regions selected were 967 to 1175 cm^{-1} and 1483 to 1771 cm^{-1}. For pH, the Selectivity Ratio algorithm selected regions all along the spectroscopic range, suggesting that pH prediction requires information from the full spectrum. Because pH has no direct effect on the molar absorptivity in MIR it is difficult to correlate it with specific regions, but it can be done with those associated with the absorbance of acids and their deprotonated forms[16]. In **Article 2**, different validation strategies (internal CV, halves splitting and Kennard Stone) were introduced to improve the robustness of PLSR models. Similar RMSECV and RMSEP were obtained (0.0014 $g \cdot mL^{-1}$ for density and 0.06-0.07 for pH), suggesting that the models were robust and that parameter prediction was not dependent on the dataset used. In contrast to other studies, in which PLSR with low RMSECV and RMSEP are obtained by selecting a high number of latent variables (in most cases, more than 10), our PLSR models only needed one or two latent variables, so having less risk of overfitting[17]. This result is specially remarkable because previous works, achievied low SEP values for glucose and fructose through external validation, but they required high numbers of latent variables, which increases the

risk of overfitting[15]. A compromise between accuracy for the predictions and avoiding the risk of overfitting is essential.

Some authors have reported the importance of temperature control during ATR-MIR measurements. Hence, when analysing food and beverage samples, fixing a constant temperature (or a range of temperatures) for ATR-MIR measurements is crucial. In our studies, our ATR-MIR instrument reached a constant temperature when reaching 63ºC, and we detected that measuring samples during the warming up of the spectrometer greatly affected the spectra, and thus the results of the models. This behaviour agrees with that found by Fayolle *et al.* who reported how temperature fluctuations could have a negative impact in the calibration and prediction of chemical parameters in wine[18].

7.4. Preprocessing and variable selection

Spectral preprocessing depends not only on the type of signal to be collected (e.g. spectrum or chromatogram), but also on the goal of the study. Hence, despite the promising results in **Articles 1 and 2** in the prediction of sugars, pH and density during alcoholic fermentation, other preprocessing combinations, including first derivatives, were tested to enhance the presence of smaller peaks, as one of our final objectives was to detect deviations due to unwanted deviations such as the production of lactic acid, which is a minor component in wine and cannot be easily detected using MIR spectra because main spectral changes are due to alcoholic fermentation **(Article 2**). In addition, a variable selection methodology was needed, based on chemical knowledge of the spectra and trial and error approach. This allowed to shift the focus towards malolactic fermentation variability.

7.5. Principal Component Analysis

The exploratory analysis (PCA) on the dataset with reduced variables (1320 to 1109 cm^{-1}) revealed a different trend in the PCA space for the contaminated fermentations with LAB (respect to time) from the Normal Operation Conditions (NOC) samples. With only three PCs, 95% of the total variability could be explained. In the PCA, the different trajectories for the contaminated fermentation (MLF) samples appeared after 100 hours for experiments 1 and 2, while for experiment 3 it was possible to qualitative see the different trajectories after 50 hours.

PCA was a valuable tool for exploring data, using a minimum preprocessing (Savitzky-Golay second order polynomial smoothing through 7 points, SNV and mean-centering). In **Article 1**, a PCA model including all data points was built, using a time-wise unfolding.

PC1 showed the same trend as sugar consumption, confirming that alcoholic fermentation could be modelled using a portable ATR-MIR spectrometer. Likewise, the same trend was seen in **Article 2**, in which the evolution of density was compared to PC1 scores trend, showing a close similarity. In both cases, by looking at the loading values, the most important region to follow the progress of alcoholic fermentation was found to be between 950-1700 cm^{-1}. These results are similar to those obtained by Buratti *et al.*, as in a PCA from fermenting wine samples, PC1 (explaining 98% of the total variability) showed that the main wavenumbers responsible of the separation of the samples were related to carbohydrates and ethanol[19]. In another study focusing on the time course of fermentation in wild ferments, it was found that more than 90% of the variation in ATR-MIR spectra was explained using four PCs. Similarly to our studies, in this case, the highest loadings in the MIR spectra were observed in the fingerprinting region, and were mostly associated with sugars and ethanol[2].

7.6. The batch-effect

In **Article 2**, new experiments under the same conditions were conducted to evaluate how the use of data from new fermentations influenced the chemometric models. Also, new concentrations for the LAB fermentations where tested, to study different moments in which the fermentations should become deviated. When introducing samples from new experiments to build the PCA model, small differences between experiments could be observed. These differences however, could not be related to any specific region of the spectra. This effect is commonly referred as "batch-effect", or even "experiment-effect". This makes it difficult to include data from other experiments into the PCA models, as no specific variables related exclusively to the experiments can be excluded, or minimized (using preprocessing strategies) from the data. Zeaiter *et al.* also observed this 'batch-effect' when analyzing NIR data arising from two different wine fermentation experiments. They suggested the application of Dynamic Orthogonal Projection for correcting NIR data and increasing robustness of chemometric models robustness[20]. In all of our fermentation experiments (**Articles 1, 2 and 3**), the consumption rate for sugars was different even though they were performed under the same experimental conditions. This poses an important problem to detect fermentation deviations, as, in some cases, the difference between batches and experiments could exceed the differences among the types of fermentation. This experiment effect is addressed in **Articles 2 and 3**. Despite the obvious usefulness of applying corrective mathematical algorithms to minimize the batch-effects present in our data, we could not implement the orthogonalization method as the use of

this type of correction would compromise the information related to the deviations we wanted to detect (sluggish fermentations and bacterial spoilage).

7.7. Classification analysis

As for the detection of deviations, discrimination from NOC fermentations was possible using PLS-DA. For the discrimination between sluggish fermentations (YAN) and NOC, the whole spectra could be used (**Article 1**). However, for the discrimination between NOC and MLF samples, a variable selection process, based on chemical knowledge of the spectra and the trial and error approach, was necessary needed **(Article 2**).

7.7.1 Sluggish fermentations

In **Article 1**, the discrimination between YAN, promoted by nitrogen deficiency, and NOC fermentations was achieved. Two types of PLS-DA discrimination models were developed, depending on the matrix arrangement: a global model and *k*-local models where *k* refers to a specific time point. Using the global model, in which data were arranged using a time-wise unfolding (including all time points), a 100% discrimination for NOC and YAN samples was achieved. Similarly, using the local strategy, where PLS-DA models were built for each individually collected sampling time, YAN and NOC samples could be discriminated with a 100% of correct classification using the models from 49 hours onwards. Comparing with other similar studies, Urtubia *et al.* found that using ATR-MIR spectra and PLSR models, glucose, fructose, ethanol, glycerol, succinic acid and acetic acid could be used as early indications of problems in nitrogen deficient fermentations. In their study, a distinctive behavior between sluggish and NOC fermentations could be seen after 30 hours[21]. Cozzolino and Curtin used ATR-MIR to predict the time course of fermentation using PLSR and obtained a coefficient of determination (R^2) and SEP of 0.93 and 1.21 days, respectively[2].

7.7.2. Lactic acid bacteria spoilage

In **Article 2**, by selecting a small window of variables, the influence of the spectral changes corresponding to the main process (alcoholic fermentation) was minimized, allowing us to distinguish between NOC and MLF samples using PLS-DA. This variable selection step clearly indicates how other important variations can mask relevant peaks when discriminating different fermentation samples. Using *k*-local models, the earliest time point in which NOC and MLF fermentations could be 100% discriminated depended on the amount of LAB inoculated at the beginning of fermentation. The fermentations with the

lowest LAB concentration were discriminated after 213 hours, while the fermentations with the highest LAB concentration could be discriminated after 56-58 hours. At those deviations times, malolactic fermentation was halfway through, giving the chance to apply corrective measures. such as the addition of sulphites, to stop the bacterial activity. We estimated that the minimum difference in L-malic acid concentration between NOC and MLF samples detected with our instrument ranged from 0.7 to 0.8 $g \cdot L^{-1}$. These results are in agreement with the literature. Fayolle *et al.* monitored lactic acid fermentation using FT-MIR in culture media; they developed PLSR models for lactose, galactose, biomass and lactic acid, obtaining a SEP value of 0.9 $g \cdot L^{-1}$ for lactic acid[22].

7.7.3. Acetic acid bacteria spoilage and ASCA

The contamination with Acetic Acid Bacteria (AAB) was achieved using an autochthonous AAB strain (**Preliminary ASCA results**). In AAB contaminations, not only was acetic acid production observed, but also a slight difference in the rate of sugar consumption between NOC and AAB fermentations. This suggests that AAB are more stressful to yeast metabolism than LAB, slowing down the alcoholic fermentation process in AAB contaminated samples. When using ASCA for exploratory analysis of NOC, AAB and MLF samples, a grouping for the contamination factor was observed after 120 hours. Three groups were found, one for each type of fermentation. These results suggested that, using ATR-MIR and chemometrics, several types of contamination can be observed during alcoholic fermentation. From the beginning of fermentation until 120 hours, the time factor was the one accounting for most of the variance in the data. During the first days of fermentation sugars are rapidly consumed, which explains why this factor retains most of the variability in the data during this period of time. Because several fermentation batches (from different experiments under the same conditions) were included for ASCA analysis, experiment effect could be seen in the data, as some part of the total variability was be attributed to this factor, especially towards the end, when alcoholic fermentation is almost finished. In contrast to the study of lactic acid bacteria spoilage, in which several preprocessing combinations had to be tested and a very small window of variables was necessary to detect the LAB contaminations, using ASCA we could observe variations among groups only using Savitzky-Golay Smoothing and SNV. This suggest that ASCA could be useful to rapidly assess if ATR-MIR data contain any related variables to the factors under study. In **Article 1**, production of acetic acid was observed in YAN fermentations. A higher production of acetic acid is commonly observed in the nutrient deficient fermentations. However, in terms of spectral data, YAN deviations were detected

because of the slower rate in sugar consumption and not because of the production of acetic acid. Thus, the first preprocessing applied (without first derivatives) was not able to detect acetic acid production. In contrast, Urtubia *et al.*[23] investigated the use of ATR-MIR data and PLSR to predict the concentration of acetic acid in wine fermenting samples and they obtained a standard error of cross-validation (SECV) of 0.18 $g \cdot L^{-1}$.

7.8. MSPC charts: global and local approaches

Deviated fermentations were also identified based on Q residuals and Hotelling T^2 statistics. In **Article 1**, using *k*-local models it was possible to distinguish between NOC and YAN fermentations with the 49-hour model, where YAN samples were out of the confidence limits for both Q and T^2 statistics. Our following step was to develop MSPC charts for fermentation monitoring because we suspected that, when describing the NOC space, abnormal fermentations would fall outside the limits of the two statistics. For all models, the reduced region, from 1320 to 1109 cm^{-1}, and preprocessing strategy including first derivatives, was required. The MSPC charts were developed at two levels: using a single experiment data or using data from different experiments.

Using samples from a single experiment, a PCA model was first developed to see if different trends could be observed between NOC and MLF samples. The first PC accounted for 98.6% of the total variability, while the second PC, showing the different trends from hours 60-80 until the end of the process, accounted only for the 1.04% of the total variability. These results are in agreement with our expectations, because the main process (alcoholic fermentation) involves the transformation of sugars into ethanol from an initial concentration of 200 $g \cdot L^{-1}$, as opposition opposed to malolactic fermentation, in which the initial concentration of malic acid is only around 2 $g \cdot L^{-1}$. This means that less than 2 $g \cdot L^{-1}$ of lactic acid will be produced, and the variability in the signal is much lower. In addition, in the fingerprinting region there are many overlapping peaks, as in this region of the infrared spectrum is where most of the organic molecules absorb (including sugars and acids). An MSPC chart was developed using only NOC samples from the single experiment. 95% confidence limits of T^2 and Q residual statistics were calculated and, when deviated fermentations were projected into these charts, they showed out-of-limit values from 80 hours onwards in the Q chart but not in the T^2 chart. From hour 110, however, T^2 values for the contaminated samples showed higher T^2 values, but they were below the confidence limit. A possible explanation for this behaviour is that malolactic fermentation, at this time, is almost finished and the distance to the centre of the model increases but not significantly,

from a statistical point of view. Some NOC samples were left out for validation purposes using the Kennard Stone algorithm. When the NOC samples were projected on the charts, all were below the limits. The fact that MLF samples are distinguishable from NOC samples in the Q-chart but not in the T^2-chart is a reasonable result. This is because the enormous variability of the spectra due to alcoholic fermentation (especially between hours 70 and 90, when the so-called tumultuous fermentation takes place) makes difficult the possibility of establishing a confidence limit to differentiate the samples due to a very small variation. Using ATR-MIR data and MSPC charts, van Sprang *et al.* monitored a polymerisation process, in which some batches were disrupted by applying a different operating temperature which affected the kinetics of the reaction. They found that abnormal batches could be detected using the T^2 chart because of the slow polymer conversion rate, but these batches were not detected in the Q charts[24]. In our study, the sugar conversion rate is always normal, which would explain why our T^2 charts are not useful to detect a bacterial spoilage where a new compound is being synthetized. In contrast, as the Q chart detects the non-modeled data, when a new or unknown compound is found in the system, it should be observed in the Q chart. Abreu *et al.* used T^2 and Q control charts to monitor the quality of green coffee beans during storage using Raman Spectroscopy. The T^2 control chart was not able to detect out-of-control samples, but the same samples appeared outside the quality control region in the Q chart[25]. In contrast, Catelani *et al.* found that deviations during the coffee roasting process could be detected using both T^2 and Q charts based on NIR data[26]. These studies suggest that the usefulness of each type of chart (T^2 and Q) should be assessed individually. Depending on the type of deviation to be detected (e.g. a conversion rate, the synthesis of unwanted compounds), the type of analytical data and the type of process in hand, the performance of the charts will vary.

When spectra of new NOC fermentation batches were included in the PCA model, these samples were below the 95% confidence limit in both Q and T^2 control charts. The Q chart built was similar to the single experiment Q chart, but for the T^2 control chart, deviated samples were above the limit from hour 120 until the end of the process. This may be explained because including more NOC samples in the PCA model increases the capability of better describing the alcoholic fermentation variability, and then it is possible to detect the differences between NOC and MLF samples, which can now be statistically differentiated, at the end of alcoholic fermentation. As described in the bibliography, when few control samples are available to build T^2 charts it becomes difficult to test whether

statistical variations between control and faulty fermentation batches are significant. A higher number of control samples can drastically improve the quality of the T^2 charts[27].

For PCA models of both single and multiple experiments, the time-wise unfolding approach was proposed as an alternative to a batch-wise unfolding. In this way, samples are time independent and the same number of sampling times is not required to project new suspected samples. In the bibliography, the batch-wise unfolding is the most common approach to monitor industrial scale fermentation data, but since batch operations have different durations, this requires the application of time synchronization methods in order to compare multiple batches. If the batches do not show small variations in duration these methods are difficult to apply[28].

7.9. Time interval-MSPC charts

T^2 values for NOC samples were high during the main stages of alcoholic fermentation (from 40 to 120 hours), and made it difficult to establish a control limit at which deviated samples could be detected. We raised the need to set control limits based on time intervals, in order to obtain more accurate confidence limits (especially for Hotelling T^2), yet no statistical separation was achieved using T^2 charts. On the other hand, the Q chart for the time interval between 65 and 72 hours showed a MLF separation trend, and could be used as an alert indicator before all MLF batches are completely separated. The complete statistical separation in the Q chart was obtained in the model between 81 and 88 hours. At the 65-72h interval, only 40-50% of this bioprocess has taken place so that readjustments could still be made in the system to avoid irreparable damage to the final wine. When using time intervals, it is possible to generate new models taking into account the variability at different points of the process, which is more representative of the current process operation than considering the whole process. Awhangbo *et al.* used the Moving Window (MW) PCA method to build MSPC charts to monitor an anaerobic digestion process using NIR, which addresses the same problem as we do with time interval models. This method "slides" along the data, to build a new fitted-model each time, including the newest sample and discarding the oldest one in each new model. They found that the global PCA model on the data was not as useful to predict anomalous samples in dynamic systems as the MW approach[29]. For our data, we preferred time intervals rather than MW because of the low number of data points, and because, in this way, different stages in alcoholic fermentation can be clearly differentiated.

7.10. Biological Process Time

In **Article 1**, the concept of biological process time applied to alcoholic fermentation was discussed. The fact that alcoholic fermentation is a biological process, a slight variation in the speed of fermentation does not necessarily mean a deviated process. Hence, it is difficult to define control limits when there are variations in the duration of fermentation. Using biological process time, both NOC and YAN spectra could be modelled against the evolving fermentation time, and the status of future batches during the process was evaluated using a PLS model. The prediction of the biological time for YAN samples confirmed that when the NOC samples are 100% fermented, YAN samples were about 60% fermented. This result showed the possibility of using biological process time as an indication of normal progression of alcoholic fermentation. Besenhard *et al.*[27] also applied the biological process time to determine the 'maturity' time of industrial fermentation batches, obtaining a good correlation for the biological time prediction and the actual response for all in-time batches.

7.11. MCR-ALS models and 'inverse' MSPC charts

In **Article 4**, the MCR-ALS resolution method was applied to extract relevant information from an experiment consisting of NOC and MLF samples. The final MCR-ALS model using the multiset structure showed good data fit (R^2 = 99,9488) and a 2,3074% lack of fit. Similarly, in a study using ATR-MIR and MCR-ALS, Grassi *et al.* monitored the alcoholic fermentation of beer. Using 6 components, their MCR-ALS model explained a 99.9% variance and had a 3.5% lack of fit[30]. In our model, four components were considered, where three of them were directly related to sugars (glucose + fructose), ethanol and lactic acid. The fourth component, however, could not be attributed to any specific species in the mixture, but it was included because it improved model performance. Because of the shape of its relative concentration curve, we suspected that this fourth component could be due to the presence of salts of tartaric acid, precipitate by the end of alcoholic fermentation. The spectral profiles found were in agreement with the literature. The lactic acid profile showed an important peak at 1150cm^{-1}, which can be attributed to the C-O stretching from carboxylic acid. As for sugars and ethanol, they appear extremely overlapped, as expected. Between 1089 and 1126 cm^{-1}, C-C and C-H stretching vibrations are found, which are very common in these organic molecules. The relative concentration profiles for sugars, ethanol and lactic acid were also appropriate. All fermentation batches showed similar concentration trends for sugars, ethanol and intermediate species. The appearance of lactic acid was only detected in contaminated fermentations. Many authors have reported the use

of vibrational spectroscopy data and MCR-ALS to model fermentation processes. Garrido *et al.* used NIR data and MCR-ALS to monitor industrial alcoholic fermentation processes. Using three components in their MCR-ALS models (sugars, ethanol and biomass), the explained variance was always greater than 99.9% and the % lack of fit smaller than 0.1% [31]. Grassi *et al.* used NIR data and MCR-ALS models to monitor milk lactic acid fermentation. Their MCR-ALS models explained 99,9% of the variance in the data and had 0.63665% lack of fit. They successfully described lactic acid fermentation and pointed the characteristic coagulation phases of milk lactic acid fementation[32]. In **Article 4**, we also investigated the use of MCR-ALS to develop MSPC charts based on the **C** matrix (relative concentrations) from the contaminated samples. These MSPC charts were used to specifically detect a fermentation deviation due to LAB spoilage. MSPC charts based on PCA models built with the relative concentration from the deviated samples were able to detect if NOC samples were out-of-limits, which would mean that the sample is not contaminated. Using the MSPC charts in the opposite way of traditional control charts means that if a projected sample falls below "control" limits, it would mean that lactic acid is being produced and thus, the model should be corrected to return to the normal conditions. Normal samples were detected in the T^2 chart after 96 hours. Oliveira *et al.* used NIR and chemometric models to detect adulteration of commercial gasoline batches based on the distillation process. Using the resolved concentration profiles from MCR-ALS decomposition of NOC samples, they built different MSPC charts. They used different PCA matrix arrangements, and they could detect all off-specifications batches both in Q and T^2 charts, using a Global-MSPC model and a Fixed size moving window -MSPC model. Using Evolving-MSPC models, some off-specification batches were not detected in the T^2 charts[33]. Finally, in another study, Grassi *et al.* were able to detect coagulation failures from the initial phases of the milk coagulation process using both T^2 and Q charts, based on NIR data for the global renneting process[34].

7.12. References

1. Pearce K, Culbert J, Cass D, Cozzolino D, Wilkinson K. Influence of Sample Storage on the Composition of Carbonated Beverages by MIR Spectroscopy. Beverages. 2016;2(4):26. doi:10.3390/beverages2040026.
2. Cozzolino D, Curtin C. The use of attenuated total reflectance as tool to monitor the time course of fermentation in wild ferments. Food Control. 2012;26:241-246. doi:10.1016/j.foodcont.2012.02.006.
3. Kölhed M, Karlberg B. Capillary electrophoretic separation of sugars in fruit juices using on-line mid infrared Fourier transform detection. Analyst. 2005;130(5):772-778. doi:10.1039/b416289g.

4. Di Egidio V, Sinelli N, Giovanelli G, Moles A, Casiraghi E. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. Eur Food Res Technol. 2010;230(6):947-955. doi:10.1007/s00217-010-1227-5.

5. Schalk R, Geoerg D, Staubach J, Raedle M, Methner F-J, Beuermann T. Evaluation of a newly developed mid-infrared sensor for real-time monitoring of yeast fermentations. J Biosci Bioeng. 2017;123:651-657. doi:10.1016/j.jbiosc.2016.12.005.

6. Preserova J, Ranc V, Milde D, Kubistova V, Stavek J. Study of phenolic profile and antioxidant activity in selected Moravian wines during winemaking process by FT-IR spectroscopy. J Food Sci Technol. 2015;52(10):6405-6414. doi:10.1007/s13197-014-1644-8.

7. Rodrigues KCS, Sonego JLS, Bernardo A, Ribeiro MPA, Cruz AJG, Badino AC. Real-Time Monitoring of Bioethanol Fermentation with Industrial Musts Using Mid-Infrared Spectroscopy. Ind Eng Chem Res. 2018;57(32):10823-10831. doi:10.1021/acs.iecr.8b01181.

8. Patz C-D, Blieke A, Ristow R, Dietrich H. Application of FT-MIR spectrometry in wine analysis. Anal Chim Acta. 2004;513:81-89. doi:10.1016/j.aca.2004.02.051.

9. Debebe A, Redi-Abshiro M, Chandravanshi BS. Non-destructive determination of ethanol levels in fermented alcoholic beverages using Fourier transform mid-infrared spectroscopy. Chem Cent J. 2017;11(1):1-8. doi:10.1186/s13065-017-0257-5.

10. Fragoso S, Aceña L, Guasch J, Mestres M, Busto O. Quantification of phenolic compounds during red winemaking using FT-MIR spectroscopy and PLS-regression. J Agric Food Chem. 2011;59(20):10795-10802. doi:10.1021/jf201973e.

11. Anjos O, Santos AJA, Estevinho LM, Caldeira I. FTIR-ATR spectroscopy applied to quality control of grape-derived spirits. Food Chem. 2016;205:28-35. doi:10.1016/j.foodchem.2016.02.128.

12. Hirri A, De Luca M, Ioele G, *et al.* Chemometric classification of citrus juices of Moroccan cultivars by infrared spectroscopy. Czech J Food Sci. 2015;33(2):137-142. doi:10.17221/284/2014-CJFS.

13. Aleixandre-Tudo JL, Nieuwoudt H, Luis Aleixandre J, Du Toit W. Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using dierent infrared spectroscopy techniques. Talanta. 2017;176:526-536. doi:10.1016/j.talanta.2017.08.065.

14. Versari A, Boulton RB, Parpinello GP. Effect of spectral pre-processing methods on the evaluation of the color components of red wines using Fourier-Transform Infrared spectroscopy. Ital J Food Sci. 2006;18:423-431.

15. Cozzolino D, Cynkar W, Shah N, Smith P. Feasibility study on the use of attenuated total reflectance mid-infrared for analysis of compositional parameters in wine. Food Res Int. 2011;44:181-186. doi:10.1016/j.foodres.2010.10.043.

16. Claßen J, Aupert F, Reardon KF, Solle D, Scheper T. Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. Anal Bioanal Chem. 2017;409(3):651-666. doi:10.1007/s00216-016-0068-x.

17. Regmi U, Palma M, Barroso CG. Direct determination of organic acids in wine and wine-derived products by Fourier transform infrared (FT-IR) spectroscopy and chemometric techniques. Anal Chim Acta. 2011;732:137-144. doi:10.1016/j.aca.2011.11.009.

18. Fayolle P, Picque D, Perret B, Latrille E, Corrieu G. Determination of Major Compounds of Alcoholic Fermentation by Middle-Infrared Spectroscopy: Study of Temperature Effects and Calibration Methods. Appl Spectrosc. 1996;50(10):1325-1330. doi:10.1366/0003702963904872.

19. Buratti S, Ballabio D, Giovanelli G, *et al.* Monitoring of alcoholic fermentation using near infrared and mid infrared spectroscopies combined with electronic nose and electronic tongue. Anal Chim Acta. 2011;697:67-74. doi:10.1016/j.aca.2011.04.020.

20. Zeaiter M, Roger JM, Bellon-Maurel V. Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. In: Chemometrics and Intelligent Laboratory Systems. ; 2006. doi:10.1016/j.chemolab.2005.06.011.

21. Urtubia A, Pérez-Correa JR, Pizarro F, Agosin E. Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. Food Control. 2008;19:382-388. doi:10.1016/j.foodcont.2007.04.017.

22. Fayolle P, Picque D, Corrieu G. Monitoring of Fermentation Processes Producing Lactic Acid Bacteria by Mid-Infrared Spectroscopy. Vol 14.; 1997:247-252. doi: /10.1016/S0924-2031(97)00004-0.

23. Urtubia A, Pérez-Correa JR, Meurens M, Agosin E. Monitoring large scale wine fermentations with infrared spectroscopy. Talanta. 2004;64:778-784. doi:10.1016/j.talanta.2004.04.005.

24. Van Sprang ENM, Ramaker HJ, Boelens HFM, *et al.* Batch process monitoring using on-line MIR spectroscopy. Analyst. 2003;128(1):98-102. doi:10.1039/b209826c.

25. Abreu GF, Borém FM, Oliveira LFC, Almeida MR, Alves APC. Raman spectroscopy: A new strategy for monitoring the quality of green coffee beans during storage. Food Chem. 2019;287(June 2018):241-248. doi:10.1016/j.foodchem.2019.02.019.

26. Catelani TA, Santos JR, Páscoa RNMJ, Pezza L, Pezza HR, Lopes JA. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. Talanta. 2018;179(October 2017):292-299. doi:10.1016/j.talanta.2017.11.010.

27. Besenhard MO, Scheibelhofer O, François K, Joksch M, Kavsek B. A multivariate process monitoring strategy and control concept for a small-scale fermenter in a PAT environment. J Intell Manuf. 2018;29(7):1501-1514. doi:10.1007/s10845-015-1192-8.

28. Mears L, Nørregård R, Stocks SM, *et al.* Multivariate Analysis of Industrial Scale Fermentation Data. Vol 37. Elsevier; 2015. doi:10.1016/B978-0-444-63577-8.50123-6.

29. Awhangbo L, Bendoula R, Roger JM, Béline F. Fault detection with moving window PCA using NIRS spectra for monitoring the anaerobic digestion process. Water Sci Technol. 2020;81(2):367-382. doi:10.2166/wst.2020.117.

30. Grassi S, Amigo JM, Lyndgaard CB, Foschino R, Casiraghi E. Assessment of the sugars and ethanol development in beer fermentation with FT-IR and multivariate curve resolution models. Food Res Int. 2014;62:602-608. doi:10.1016/j.foodres.2014.03.058.

31. Garrido M, Rius FX, Larrechi MS. Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. Anal Bioanal Chem. 2008;390(8):2059-2066. doi:10.1007/s00216-008-1955-6.

32. Grassi S, Alamprese C, Bono V, Casiraghi E, Amigo JM. Modelling Milk Lactic Acid Fermentation Using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS). Food Bioprocess Technol. 2014;7(6):1819-1829. doi:10.1007/s11947-013-1189-2.

33. de Oliveira RR, Pedroza RHP, Sousa AO, Lima KMG, de Juan A. Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy. Anal Chim Acta. 2017;985:41-53. doi:10.1016/j.aca.2017.07.038.

34. Grassi S, Strani L, Casiraghi E, Alamprese C. Control and monitoring of milk renneting using FT-NIR spectroscopy as a process analytical technology tool. Foods. 2019;8(9). doi:10.3390/foods8090405.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Chapter 8

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

Conclusions

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

In this thesis, monitoring of alcoholic fermentation using a portable ATR-MIR instrument and chemometric techniques has been investigated. The main conclusions arising from this work are:

- A portable ATR-MIR spectrometer coupled with PLSR can quickly monitor sugars, pH and density in fermenting samples during all the stages of wine alcoholic fermentation, with no need of taking samples to a laboratory. Hence, it is a rapid and economical strategy to monitor fermentation processes.

- Fermentation problems can be detected using ATR-MIR spectra and chemometric models, giving the winemaker the opportunity to eventually correct the process and obtain a good quality product.
    - Sluggish fermentations can be detected at an early stage of fermentation using multivariate discrimination techniques.
    - Lactic acid bacteria spoilage can be detected before the end of malolactic fermentation, that is, before total conversion of L-malic acid into lactic acid.

- Q residuals and Hotelling T^2 statistics offer the possibility to develop MSPC charts based on the information from Normal Operation Conditions samples, and detect abnormal samples during alcoholic fermentation.
    - In the T^2 control charts, the slower kinetics of sugars conversion into ethanol is detected as abnormal, and can be used to detect sluggish fermentations.
    - Q charts can detect deviations during wine alcoholic involving the apparition of undesired molecules. Because lactic acid is produced during lactic acid bacteria spoilage, Q charts can be used to detect its presence.

- MSPC charts are a valuable tool to monitor wine alcoholic fermentation. Time interval-MSPC charts have better performance than charts based on global models, where all the variability is considered in a single model.

- ASCA can assess the statistical significance of minor variations due to experimental design factors, such as bacteria spoilage. The application of ASCA before the development of any chemometric model can optimise the selection of variables that are influencing a factor of interest.

- ATR-MIR coupled to MCR-ALS can provide important information to evaluate alcoholic fermentation progression.
    - MSPC charts based on MCR-ALS data allow the detection of contaminated samples due to the production of lactic acid bacteria in wine fermenting samples.

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT ROVIRA I VIRGILI
MONITORING WINE FERMENTATION
USING ATR-MIR SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES
Julieta Cavaglia Pietro

UNIVERSITAT
ROVIRA i VIRGILI