QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE

Daniel Schorn García

UNIVERSITAT ROVIRA i VIRGILI

QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE

Daniel Schorn-García

PhD. THESIS

2023

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

DANIEL SCHORN GARCÍA

Quantifying Variability in Grape and Wine Quality: A Multivariate Analysis Perspective

Doctoral Thesis

Supervised by Ph.D. Montserrat Mestres Solé, Ph.D. Ricard Boqué Martí and Ph.D. Barbara Giussani

Department of Analytical Chemistry and Organic Chemistry
Universitat Rovira i Virgili

UNIVERSITAT ROVIRA i VIRGILI

Tarragona, 2023

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

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

Ph.D. Montserrat Mestres Solé, Aggregate Professor of Analytical Chemistry and Ph.D. Ricard Boqué Martí, Associate Professor of Analytical Chemistry at the Department of Analytical and Organic Chemistry of Universitat Rovira i Virgili, and Ph.D. Barbara Guissani, Associate Professor of Analytical Chemistry at the Department of Science and High Technology of University of Insubria,

CERTIFY,

that the Doctoral Thesis entitled "**Quantifying Variability in Grape and Wine Quality: A Multivariate Analysis Perspective**", submitted by Daniel Schorn García 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 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, 30th June 2023,

Ph.D. Montserrat Mestres Solé	Ph.D. Ricard Boqué Martí	Ph.D. Barbara Guissani

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

This work has meant the consolidation of one of the research lines of the groups Instrumental Sensometry (iSens) and Chemometrics, Qualimetry and Nanosensors (QQiN), resulting in the creation of the Chemometrics and Sensorics for Analytical Solutions (ChemoSens) research group of the Universitat Rovira i Virgili.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

This work constitutes a part of the research developed by the Instrumental Sensometry Group of the University Rovira i Virgili. All the results here presented have been obtained in the iSens laboratory at the Faculty of Oenology of Tarragona.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

This work constitutes a part of the grant PID2019-104269RR-C33 funded by the Spanish Ministry of Science and Innovation and the State Research Agency (MCIN/AEI/10.13039/501100011033).

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

The present PhD thesis has been possible thanks to the financial support given to Daniel Schorn García (2020 FISDU 00221 grant) by the *Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya*.

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

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

"*It is through science that we prove,*
but through intuition that we discover"

- Henri Poincaré -

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

A mi madre,
a mi abuela

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Agradecimientos

Ha llegado el final de una etapa muy importante en mi vida. La realización de esta tesis doctoral ha sido un camino lleno de experimentos, horas en el laboratorio, experiencias y muchas personas. Las primeras personas que me vienen a la mente, y a las que tengo que agradecer que esta tesis sea tal y como es, son mis tres supervisores (mis tres "jefes") Montse, Bárbara y Ricard. En primer lugar, quiero agradecer a Montse, por darme la oportunidad de empezar en este grupo desde el máster, por tu positividad y tus palabras de ánimo, tu confianza en mí, tu increíble capacidad científica y por saber ponernos límites cuando diseñábamos el experimento. **Gràcies, Montse**. A Barbara, por tantísimas horas enseñándome quimiometría, por hacer que la distancia a Italia fuera cortísima, por sus increíbles ideas y por todos los ratos compartidos, sobre todo los acompañados de un buen vino. **Grazie, Bárbara**. A Ricard, por tener siempre la corrección justa, por ayudarme en todo lo que le pedía, por tener esa serenidad que tanto hacía falta en algunos momentos y por los chistes malos que sentaban tan bien. **Gràcies, Ricard**. Ha sido un placer tenerles a mi lado en esta experiencia.

He de decir que el apoyo de los "seniors" no solo acaba en Montse, Barbara y Ricard. A Laura, tengo que agradecer haberme dejado ver a una gran docente en acción en las prácticas de Análisis Instrumental, sus "vamos, que ya lo tienes" y tener siempre siempre un momento para una duda, explicación o cualquier otra cosa. **Gracias, Laura**. No puedo olvidarme por supuesto, de la "jefa suprema". A Olga quiero agradecer su capacidad de transmitir la importancia de la divulgación, su apoyo constante, sus palabras de ánimo y el explicarme como funciona una universidad y una facultad. **Gracias, Olga**. No me puedo olvidar de Santi y Jaume, con los temas que parecía que había que preparar para hablar en el café, por ayudarme a programar, a arreglar figuras, a cualquier cosa que se pudiera decir "pregúntale a Santi o Jaume que seguro que sabe". **Gracias, Santi**. **Gracias, Jaume**. He de decir que, de la Facultad, y aunque en Constantí, no puedo olvidarme de agradecer la inmensa ayuda de Pedro y todo el personal de la finca, gracias por ayudarme en las vendimias, explicarme cualquier concepto, dejarme catar e incluso colaborar en la divulgación que organizaba. **Gracias, Pedro**.

Si en Facultad tengo que agradecer, en el laboratorio, no sé que tengo que decir. A mis dos apoyos más incondicionales, mis amigos y mis compañeros de tantas horas, recordando cualquier referencia televisiva española. A Ángel tengo que agradecer su apoyo, su bondad, su preocupación, su disposición, sus preguntas indiscretas, sus tarareos mientras integraba y ser una persona increíble. **Gracias, Ángel**. Solo puedo decir ojalá hubieras llegado antes Jokin, gracias por ser increíblemente bueno conmigo y con mis líneas de código. **Mila esker, Jokin**. No me puedo olvidar de los que han pasado y han dejado su huella. **Gracias, Julieta, Aziz y Carolina**. Finalmente, quiero agradecer los granitos de arena de esta tesis de Jordi, Helena y

Nùria. Muchas gracias por vuestro trabajo. **Gràcies, Jordi** i **Helena**. A Nùria especialmente en esta etapa de redacción de tesis por sus abrazos, contarme cotilleos y traerme chocolate. **Gràcies, Nùria**.

Me gustaría agradecer a José Luis, la oportunidad de disfrutar una experiencia vital y científica como han sido esos tres meses en Sudáfrica. Estoy muy agradecido de aprender tanto sobre el análisis de compuestos fenólicos, involucrarme en el día a día del laboratorio y tener la confianza para poder volver a Sudáfrica. Gracias Jose Luis. To my shared braincell labmate, Kiera, I want to thank you so much to welcome me so well, so friendly and so warm. I do not know how could I survive so many hours in the lab without singing and talking about everything with you. **Thank you, enkosi, Dankie, Kiera**. I want to extend my gratitude to the rest of the SAWRI and especially the master student from Enolab. Me gustaría aprovechar para agradecer a Alba ser mi red de seguridad en Sudáfrica, encontrarte en la otra punta del mundo ha sido muy especial. **Gracias, Alba**.

Mi periodo en Tarragona empezó haciendo el máster, y eso fue una gran oportunidad de conocer gente maravillosa, con la cual estoy muy contento de seguir contando después de haber acabado. A mi wey, muchas gracias por tu felicidad contagiosa. **Gracias, Luis**. A mi compañero incansable de piso, en aquel piso tan loco y que se convirtió en una experiencia maravillosa. **Gracias, DaniZ**. No se puede hacer un máster sin tener un poco de safreig, a Julia por ser tan feliz en la vida. **Gràcies, Julia**.

Cuando llegué al laboratorio parecía que la tercera planta estaba muy lejos de Bioquímica, pero conseguimos que eso dejara de ser así. Mucha gente ha pasado, pero las personas con las que compartes croquetas hechas en una cadena de montaje, ves con tanta pasión Eurovisión y quemas el Apple de tanto ir, tienen siempre un trozo de tu corazón. **Gracias, Miguel y Chema**. Hago extensible todo lo anterior, a otros dos croquetos, a Mari agradezco su duliciosa forma de enseñarme a divulgar la ciencia. **Gracias, Mari**. A Aitor, porque gracias a ti he crecido como persona y por tu cariño. **Gracias, Aitor**. No me puedo olvidar de otras personas que me han hecho sentir super bienvenido abajo. **Gracias Mercé, Helena y Paloma**. A la persona más flamenca que te puedes echar a la cara, al lunar de un traje de la feria hecha persona, tu felicidad es contagiosa. **Gracias, Marta**. Al resto de personitas que de una forma y otra me han acompañado comiendo, tomando café y más. **Gracias**.

Tarragona también me ha traído un núcleo de amigos, que son maravillosos, totalmente diferentes entre ellos y eso es lo genial. Nadie puede tener una mejor compañera de películas de Marvel, me encanta tu felicidad y tus 9,5. **Gracias, Carla**. A una persona maravillosa, mi compañera de viaje al otro lado del charco, gracias por escuchar tan bien y por los cotilleos que me cuentas. **Gracias, Sandra**. A la mejor enfermera de mi CAP de referencia, una persona

fiestera y loca y además increíblemente buena. **Gracias, Eva**. Tener conversaciones profundas sobre feminismo, racismo y demás y que después un abrazo lo arregle todo es maravilloso. **Gracias, Violeta**. No puedo dejar de agradecer a la persona que me enseñó Tarragona, me enseño la catalanidad y a la vez también la esencia del sur, por todos los momentos que vivimos. **Gracias, Raúl**.

Hablar del hogar es difícil, pero yo tengo uno genial. Mi hogar no es tanto esas cuatro paredes que contienen el caos en el que hemos vivido, mi hogar es esa persona tan buena, tan maravillosa y porque no decirlo, tan nubli. A Candela tengo que agradecerle primero el aguantarme, lo sé no ha sido fácil. Me ha encantado contar con un apoyo tan incondicional, tan de hermana como el tuyo. No puedo expresar todo lo bueno que has hecho por mí. **Muchas gracias, Candela**. Agradecimiento que hago extensivo a Raquel, por ser como una madre cuando venías de visita. Al resto de los nublis, Dani y Ginger, también tengo que decir que les agradezco que me aguanten, que no ha sido fácil. He sentido mucha felicidad de poder tenerlos en esta etapa de mi vida. **Gracias, Dani y Ginger**.

Los de toda la vida han estado ahí, quizás no llegaban a saber qué hacía tantas horas en el laboratorio, o se perdían sobre los artículos que escribía, pero que sentían una alegría enorme con cada buena noticia y que eran un apoyo totalmente incondicional. **Gracias, Rubén, Rebeca, Iris, Laura y Cristina**. A mi par de amigas o a mi par de enlaces no enlazantes, gracias por estar siempre ahí. **Gracias, Pau y Clau**. A Bea, por ser esa persona loca, y a la vez tan buena. **Gracias, Bea**. A la recién llegada a Tarragona, me encanta tu forma de ser y todo el cariño que me has dado, quiero agradecerte todas las veces que de repente te acordabas de mí. **Gracias, Ana**. La etapa de pasar por la Universidad de La Laguna me trajo aún más gente maravillosa. A Idaira y Adrián tengo que agradecer el enseñarme cómo era hacer un doctorado a la vez que ser tan buenos conmigo. A mi director de TFG por sus preguntas y su manera tan didáctica de enseñarme. **Gracias, Adrián**. Aunque tarde en escuchar los audios, y aunque nos separen tantos kilómetros, siempre ha estado ahí. **Gracias, Idaira**. Finalmente, a mi gallega favorita, porque por nada del mundo se rinde, por nada del mundo deja de ser feliz, es una inspiración. **Gracias, Uxia**.

Aún habiendo agradecido a todas las personas que han estado antes y durante el camino, hay dos personas sin las cuales no habría llegado aquí. A mi madre, por intentar entender todo lo posible de qué iba mi tesis, has sido un gran apoyo y todo un referente para mí. **Muchísimas gracias, Marisol**. A mi abuela agradecer la didáctica que he conseguido gracias a tener que explicarle tantas veces que hacía y porque no iba a clase, aunque realmente por su apoyo y cariño en cualquier momento. **Gracias, Nieves**. Al resto de mi familia, agradezco su apoyo, su interés y sus preguntas enológicas en cualquier momento.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Index

Introduction

1. Grape and Wine

Wine has been an integral part of human civilization since its beginnings, and its importance is reflected in the fact that significant historical agreements and events have been witnessed by it. Wine production can be traced back to the Neolithic period, being its birthplace the region encompassing Georgia, Armenia, and Iran. From there, its consumption spread to Egypt and eventually throughout Eurasia. Greece and Rome played a vital role in the development of wine production and technology, which helped to expand the wine industry further. However, the widespread of viticulture occurred with the spread of Christianity, with monasteries leading the way in wine production[1].

Grapevine cultivation and wine consumption in Spain date back to the third millennium BC, with vine cuttings imported into the Iberian Peninsula by the Phoenicians, who brought knowledge of viticulture and winemaking to Southern Spain[2]. Wine and spirits continue to be essential elements of Spanish culture and tradition. Wine production has a major impact in Spanish economy as it represents 1.5% of the gross value added[3]. Spain stands out in wine production not only in economic terms, but also in terms of vineyard area, being the country with more surface area and more organic vineyards[4,5]. In addition, together with Italy and France, Spain is positioned on the podium in wine production, in terms of volume, with these three countries accounting for half of the world's total wine production[6].

1.1. Viticulture

The vineyard is where the process of producing wine starts and it all begins with the cultivation of grape berries. The scientific name for grapevine is *Vitis vinifera* L., and it belongs to the Vitaceae family of plants[7]. Grapevine is a widely cultivated fruit crop that provides berries used as fresh fruit, raisins but also to be fermented and distilled to produce alcoholic beverages. The genus Vitis is considered the most important genus in agronomic sciences, with Vitis vinifera being the most widely cultivated and the only species used in the global wine industry.

Vitis vinifera L. includes two subspecies: *V. vinifera* L. spp *vinifera* and *V. vinifera* L. ssp *sylvestris*. The latter is the wild form, considered the ancestor of the *V. vinifera* spp *vinifera* subspecies, which is the cultivated form. While their morphological differences may have

resulted from human domestication rather than geographical isolation, cultivated grapevines with uniform and stable characteristics are known as "varieties" or "cultivars". Domestication of grapes involved several changes in their biology and morphology to ensure better fermentation and more regular wine production[7,8]. Today, there are thousands of *Vitis vinifera* cultivars generated by vegetative propagation and crosses, but wild-type cultivars are rare. However, the global market for grape production is dominated by only a few cultivars that are classified according to their final production: wine grapes, table grapes and raisins[7].

1.2. Grape maturity

Grape maturity encompasses various aspects that contribute to both grape and wine quality. It can be understood as the culmination of three distinct maturities, each influencing different aspects of the grape and subsequently, the resulting wine.

The first is the technological maturity, which relates to the progressive development and accumulation of fundamental chemical compounds like carbohydrates and acids. These components play a crucial role in determining the taste and overall composition of grapes and wine[9].

The second is the phenolic maturity, which holds particular significance in red grape cultivars. Phenolic compounds, responsible for colour, mouthfeel and stability of the wine, evolve and accumulate during this stage. However, when dealing with phenolic maturity, polyphenolic concentration is considered as well as the possibility to extract them during winemaking. Achieving optimal phenolic maturity is essential in producing red wines with desirable organoleptic characteristics and ensuring their longevity[9].

Lastly, we have aromatic maturity, responsible for the primary aromas found in wine. These aromas originate directly from the grapes themselves or from aromatic precursors that are released during the fermentation process. Aromatic maturity is a crucial factor in the overall sensory experience of the wine, as it contributes to its distinctive bouquet and aromatic profile[9].

Understanding and managing these three interconnected maturities is vital for winemakers, as it allows them to make data-based decisions regarding the ideal harvest

time and ensures the production of high-quality wines. Winemakers can unlock the full potential of their grapes and create wines that truly showcase their unique terroir and varietal characteristics by carefully considering technological, phenolic, and aromatic maturities[9].

1.2.1. Sugars

The most remarkable and significant process in the development of berries is the accumulation of sugars inside the fruit. These sugars primarily come from the photosynthesis process. While the leaves are the most productive and efficient photosynthetic organs, it is important to note that all green organs, including the fruit before veraison, can produce sugars[9]. Carbohydrates produced through photosynthesis include glucose, fructose, sucrose, and starch. Starch remains as a stable reserve form and undergoes little change during the growth period. Glucose and fructose are initially used by the leaves themselves, and the remaining sugars are transported to other parts of the plant in the form of sucrose through the phloem with the priority of the carbohydrates receptor in the vine varying depending on the plant's vegetative cycle[10]. During the early growth phase of the berry, sucrose is supplied from mature leaves and converted into glucose and fructose within the plant cells. However, the increase in sugar concentration in the berries is mild during this phase because the sugars imported are partially used for fruit development, especially seed growth and ripening process[11].

During the initial ripening phase, a substantial accumulation of glucose and fructose in the vacuoles of the pulp cells is achieved. However, the rate of accumulation decreases as the ripening period progresses, with a peak at full maturity, which can contain in that phase between 150 to 250 $g \cdot L^{-1}$ of glucose and fructose, with a ratio between hexoses of about 1.0[12].

1.2.2. Acids

The total acidity of grapes is mainly determined by the concentrations of tartaric acid and malic acid, which make up approximately 90% of the acids found in grapes. Other acids derived from grape metabolism are also found in grapes in a smaller concentration, such as succinic or citric acids[13].

Tartaric acid, specifically the L-tartaric enantiomer, is accumulated in grape berries as a secondary product of sugar metabolism. During the growth phase, tartaric acid rapidly accumulates in the berries, reaching levels of up to 20 $g \cdot L^{-1}$. After veraison, the concentration of tartaric acid slightly decreases and remains relatively constant during the ripening period, between 10.0 and 3.5 $g \cdot L^{-1}$. The amount of tartaric acid in grapes is directly related to environmental humidity and inversely related to temperature[11].

Malic acid, primarily consisting of the L-malic enantiomer, accumulates in the vacuoles of the berry during the growth phase. This accumulation occurs as a by-product of sugar breakdown, although there is also a route involving the beta-carboxylation of pyruvic acid[14]. Starting from veraison, the concentration of malic acid begins to decrease. It is used as a substrate for cellular respiration and can also be used in the biosynthesis of glucose through gluconeogenesis. However, the amount of malic acid converted to glucose is small, accounting for less than 5% of the total accumulated malic acid. The concentration of malic acid ranges between approximately 1 to 5 $g \cdot L^{-1}$ at the time of harvest[9].

Both tartaric and malic acids, which determine the total acidity and pH of the grapes, also act as ionic regulators during berry development. The acids accumulate in cell vacuoles, but their acidity decreases from veraison onwards due to malic acid combustion during respiration, neutralization by imported cations, and dilution caused by water accumulation[11].

Despite that sugars and acids play a significant role in the technological maturity of grapes and in wine primary characteristics, without an optimal phenolic and aromatic maturities grapes do not release their full potential.

1.2.3. Polyphenolic compounds

Phenolic compounds play a vital role in influencing the organoleptic properties of grapes, giving them their characteristic colour and mouthfeel sensations[15]. These compounds exhibit an increase in concentration as grapes mature, although the relationship is not strictly linear. Among the various polyphenolic compounds, a division is made between non-flavonoids, present in the flesh of the grape, being the hydroxycinnamic acids the predominant ones, and flavonoids, such as anthocyanins and tannins, present in the skin and seeds of the grape. They are of utmost importance due to their influence on colour and

mouthfeel, respectively. However, their evolution and distribution vary within different grape parts[16].

Anthocyanins primarily reside in the grape skin[17], and their concentration intensifies as maturity progresses, reaching a peak around the technological maturity stage before eventually declining. On the other hand, tannins exhibit different behaviour depending on the grape part. Tannins from stems and seeds maintain a consistent concentration, even a slightly decrease due to polymerization, from veraison onwards[18]. Conversely, tannins derived from the grape skin experience an increase in concentration, with the highest levels achieved at the technological maturity stage[19].

It is worth noting that the convergence of technological and phenolic maturity in grapes is being impacted by climate change. Grapes are now achieving their technological maturity earlier than twenty years ago, as in the beginning of this century the first were the first reports of changes in viticulture due to climate change. This maturity pattern change is leading to a shift in the correlation between these two maturation stages[20]. This changing pattern underscores the dynamic nature of grape development under actual climate change conditions[21,22].

1.2.4. Aroma

Secondary metabolites, such as aroma and aroma precursor (aroma in a glycosylated form[23]) compounds are accumulated in grapes at the later stages of maturity. They are present in flesh and skins of grapes. Many odorant compounds present in grapes play a significant role in varietal aroma and will remain after the winemaking process[9]. The key families of aroma include terpenoids (monoterpenoids, sesquiterpenoids, and C_{13} norisoprenoids), as well as various precursors of aromatic aldehydes, esters, and thiols. Grape berries lack cellular structures for storing lipophilic volatile organic compounds, which are instead stored as water-soluble glycosides or conjugates with amino acids such as cysteine[9]. Enzymes called glycosidases and peptidases play a crucial role in releasing volatile aroma compounds from these water-soluble forms during alcoholic fermentation[24].

Although the accumulation of these compounds does not seem to be directly correlated with technological maturity, it is essential to pay close attention to their development, due to actual and future discrepancies related to climate change[22]. It has been demonstrated

that in the actual context, the concentration of typical aromas of specific cultivars, would change inversely to temperature[20].

1.3. Grape analysis

As mentioned in the previous section, determining major grape compounds to evaluate the overall maturity is of utmost importance. Traditionally, viticulturists employ a representative sampling method, collecting between one and two hundred grapes to take into account the field's variability accurately[9,25]. Thus, to ensure the results reliability, the random sampling performed should take into account several sample characteristics, such as position in the field (avoiding external vines), side of the row, position of the bunch and position of the grape within the bunch[26]. Several parameters are then determined in the sampled grapes, including:

- Average berry weight: By weighing a specific number of grape berries, usually a hundred, and calculating the average weight per berry[26].
- Sugars: The grape must obtained from the collected grapes is analysed by refractometry or by densimetry, and the concentration of sugars is expressed as $g \cdot L^{-1}$ through conversions using reference tables[27].
- pH: Determined by potentiometry using a pH-meter[28].
- Total acidity: Determined by volumetric titration of the grape must using a standardized base, usually NaOH[27].

These parameters provide valuable insights into the technological maturity of grapes as they offer information regarding sugar and acid content and, if the grapes will be fermented, determination of Yeast Assimilable Nitrogen (YAN) is also performed usually by Sorensen method in wineries, or in external laboratories by enzymatic analysis[9,29].

On the other hand, to assess polyphenolic and aromatic maturities there is not a widespread and easy-to-use methodology that can be applied during harvest, although there are some methods to evaluate their concentration and extractability[30–32]. However, each method is specific due to differences in the extraction methodology and may be used just for comparative purposes.

Currently, viticulturists mainly rely on organoleptic inspection, primarily focusing on aromatic characteristics as well as colour, hardness and mouthfeel of the grape skins and seeds[13,25]. This methodology depends on the experience of the viticulturist; however, several processes occurred throughout the maturation process may be taken into account to perform the organoleptic evaluation[25]:

- Stunted pedicel and grape grains increase their sensitivity to seeding.
- Skin becomes more vibrant, with yellow (white varieties) or bluish black (red varieties) tones. Ripe grapes easily separate from skin and stain fingers. Immature skin is acidic with low astringency, while ripe grape skin is fruity with balanced tannins.
- Pulp starts acidic and herbaceous, turning sweet and fruity[9].
- Seeds become brown and brittle, lose herbaceous flavours and may have toasted aromas when bitten.

1.4. Winemaking process

Winemaking involves a series of processes to transform grape must into wine and both technology and timing of its application influence these processes[33]. The winemaking steps are the ones explained below, although it should be noted that sometimes some of these steps can be omitted or the order can be changed.

- Destemming: Separating the woody parts of the grape bunch, such as stems, leaves, and other debris, enhances the must quality by avoiding astringent and vegetal characters while reducing oxidation and improving finesse and roundness.
- Crushing: Breaking the grape skin to release the must while being careful not to break the seeds, which could lead to bitterness and astringency in the must.
- Must clarification: Usually grape must is maintained approximately one day at low temperature to precipitate some solids and have a clearer liquid to ferment. It is performed in white wine fermentation.
- Maceration: When the must comes into contact with grape skins, usually at low temperatures, it acquires colour, mouthfeel and aroma, which contribute to the structure and sensory qualities of the wine. Maceration duration varies depending

on the type of wine, with longer durations for red wines, shorter durations or none for white wines (to prevent fermentation), and moderate durations for rosé wines.

- Alcoholic fermentation: This is the main process in wine production and implies the biochemical transformation of sugars into ethanol and carbon dioxide (CO_2) by yeast, mainly *Saccharomyces cerevisiae*. In fact, as it is defined by the International Organisation of Vine and Wine (OIV), wine is the product obtained exclusively by a total or partial alcoholic fermentation of fresh grapes, crushed or not, or of grape must. In addition to the transformation of glucose and fructose to ethanol, yeast also incorporate sugars into other metabolic pathways, which will result in the release of many metabolites[34].

- Malolactic fermentation: In red wine production, and sometimes in white and rosé wines, malolactic fermentation occurs after alcoholic fermentation so it is usually called the second fermentation. This fermentation is essential for aging red wines, while in white wines it is used selectively to soften acidity because it is the biochemical transformation where sour L-malic acid converts to softer L-lactic acid by lactic acid bacteria (LAB)[35]. This process reduces wine acidity, enhances aroma and complexity, and results in a smoother and fuller-bodied wine. However, it can reduce fruitiness and lead to undesirable odours like excessive buttery or lactic notes[36].

- Devatting: It involves separating the wine from the grape skins, seeds and stems. This is achieved by transferring the wine to another tank through racking.

- Pressing: This step plays a crucial role in the final wine quality as it extracts the remaining wine in pomace through pressure. Depending on the applied pressure, different types of wine with different chemical compositions are obtained. Higher pressure results in greater mechanical extraction from seeds and skins, but it can negatively affect the chemical and sensory qualities of wine.

- Aging: The wine obtained from the previous steps is transferred to oak barrels for this purpose, typically oak barrels but several studies have tested other woods[37]. Oak wood is carefully selected based on its hardness, permeability, porosity and also its origin, typically French or American. Moreover, the barrels are chosen according to the toasting level of the wood and the number of times they have been used over time as both parameters will influence the wine's character. This process involves various physical and chemical changes of wine facilitated by micro-

oxygenation through the barrel. The controlled exposure to oxygen promotes the polymerization and interaction between anthocyanins and tannins, resulting in both colour stabilization and astringency decrease. In addition, it should be noted that oak provides a series of volatile compounds (all of them belonging to the group of tertiary aromas) both naturally present in the wood and formed during the wine's evolution in the barrel[34]. Other materials have been explored due to its properties, such as clay and concrete[38].

- Finning: After fermentation and/or aging, it is essential to clarify, stabilize, and filter the wine to eliminate possible microbiological or clarity problems, ensuring the production of a high-quality wine.

Figures 1 and 2 display schematic representations for white and red winemaking process, respectively, which shows the slight differences that exist between both processes.

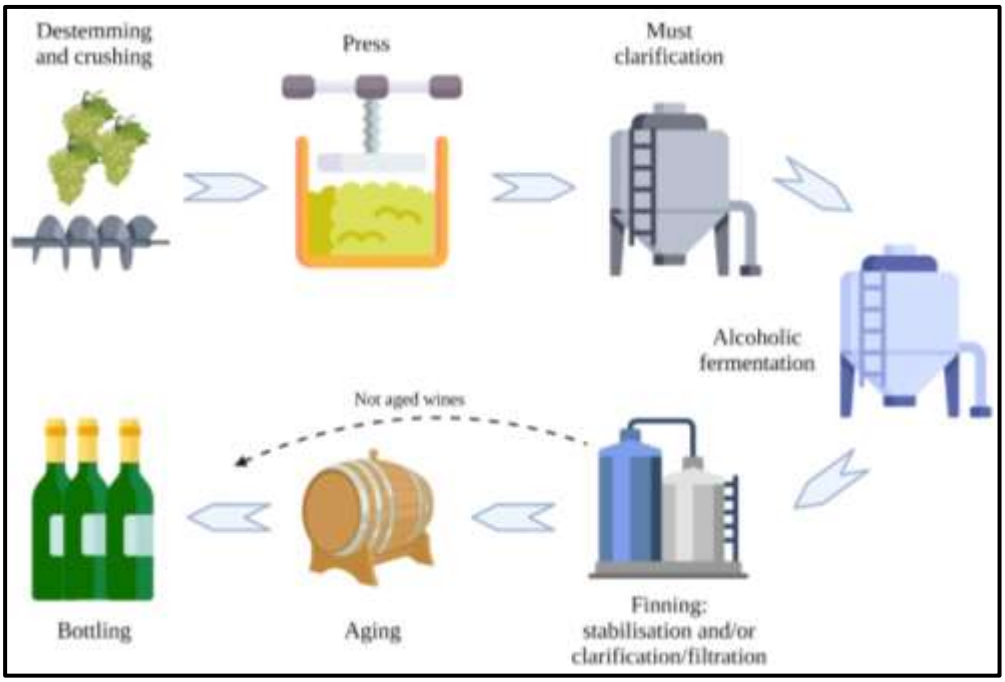


Figure 1. Main steps of white vinification process.

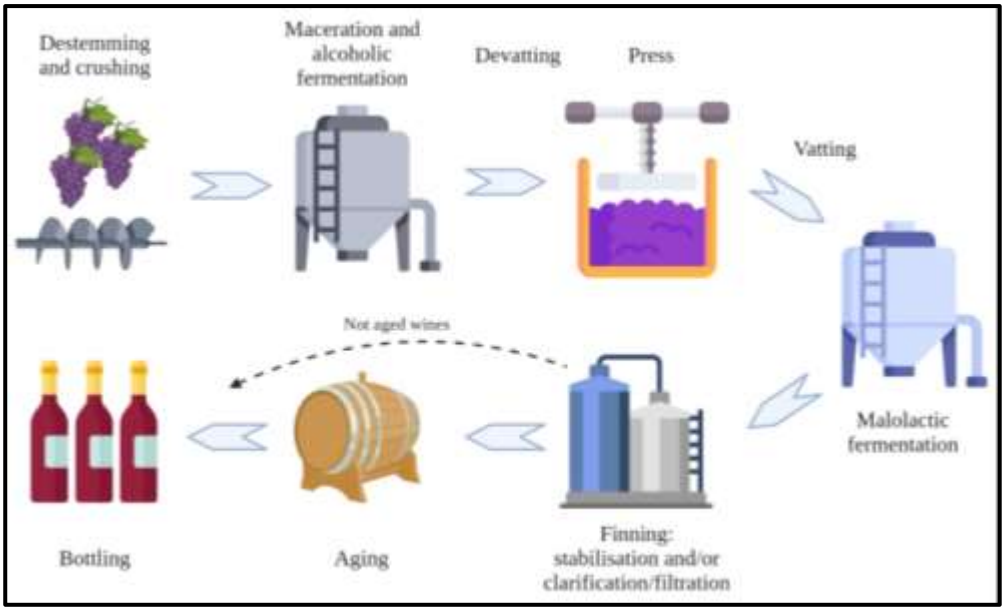


Figure 2. Main steps of red vinification process

1.5. Microorganisms

Microorganisms play a crucial role throughout the winemaking process. Some of them come from the field and are therefore naturally found in grapes but others are consciously added at specific times to perform specific functions.

1.5.1. Yeast

Among the different microorganisms found in grape skins, yeasts are the predominant with populations ranging from 10^4 to 10^6 colony-forming units (CFU) per gram of grape. It contains a diverse array of yeast species with non-*Saccharomyces* yeasts being prevalent, which would also be predominant in the early stages of alcoholic fermentation. As the fermentation progresses, the yeast population changes due to the different ability of genera to withstand high-stress conditions. As *Saccharomyces cerevisiae* presents high capacity for fermentation and resistance to ethanol it becomes the dominant yeast from the middle stages of fermentation onward, although some wineries opt to introduce this yeast earlier through inoculation[39].

1.5.2. Lactic Acid Bacteria (LAB)

LAB are responsible for the biochemical transformation of L-malic acid into L-lactic acid, known as malolactic fermentation. In grapes and fresh grape must, LAB are found in relatively low concentrations (around 10^2 to 10^3 CFU $\cdot$ mL^{-1}). The main species include *Lactiplantibacillus plantarum* and *Oenococcus oeni*[40], with the latter being more prominent during alcoholic fermentation due to its adaptation to high ethanol content. During malolactic fermentation, the decarboxylation of a dicarboxylic acid releases CO_2, resulting in a decrease in acidity and an increase in pH. Similar to yeast, LAB also produce other metabolites that affect the organoleptic properties of wine[41].

1.5.3. Acetic Acid Bacteria (AAB)

Acetic acid bacteria (AAB) are responsible for the biochemical conversion of ethanol into acetic acid, a process known as acetification. Although present in grapes at a relatively low population (10^2 and 10^3 CFU $\cdot$ g^{-1}), their presence can impact the quality and organoleptic properties of the final wine. The anaerobic environment created during alcoholic fermentation helps keep the AAB population under control. Moreover, the challenging conditions of the later stages of alcoholic fermentation make it difficult for AAB to thrive but it is not an impossible process, so its proliferation must be controlled. In fact, in the vinegar production process, these bacteria are intentionally utilized to convert wine into vinegar, because their high adaptation to high ethanol media.

1.6. Factors affecting alcoholic fermentation

The primary role of yeast is to metabolize sugars, specifically glucose and fructose, as the carbon source for the Krebs cycle. This metabolic pathway leads to the production of ethanol, as yeast undergoes partial Krebs cycle, due to the high sugar concentration which triggers the Crabtree effect[42]. However, many compounds are also necessary to ensure a proper function of the Krebs cycle and other metabolic processes in yeast.

Nitrogen compounds are of outmost importance as they are required for protein and enzyme synthesis. Nitrogen compounds can be classified as inorganic (ammonia salts) or organic (such as amino acids or peptides)[13]. However, the most interesting from an oenological point of view is the so-called YAN, which encompasses all nitrogen sources

except proline, which requires oxygen for assimilation[43], and proteins. To ensure a successful fermentation process, it is crucial to maintain an optimal YAN concentration ranging between 140 to 400 mg $N \cdot L^{-1}$. A YAN concentration below this range would hinder proper fermentation, while exceeding it could lead to the synthesis of undesirable aromas[44,45].

In addition to YAN, other micronutrients such as minerals, vitamins, and anaerobiosis factors (compounds that cannot be synthesized during fermentation due to lack of oxygen) are essential for the successful progression of alcoholic fermentation[13,45]. While these compounds play a crucial role, it is worth noting that grape must often contains an adequate concentration of these nutrients. Nonetheless, when YAN additions are carried out, they can also serve as a manner of introducing other necessary compounds. This is because commercially available supplements typically include a combination of these essential nutrients alongside YAN[46].

Several other factors play a significant role in ensuring proper yeast performance. Temperature and sulphur dioxide are two of them. Yeast operates within a specific temperature range that directly affects the integrity and permeability of its cellular membrane. The optimal temperature range for yeast activity is usually between 16 and 30 ºC[13]. However, during fermentation, the presence of ethanol requires a lower temperature to prevent its entry into the intracellular environment[47]. Proper temperature control is crucial for achieving high-quality wines and ensuring a successful alcoholic fermentation process. By carefully managing the appropriate temperature throughout fermentation, winemakers can optimize yeast performance, maintain cellular integrity and facilitate the desired metabolic pathways to produce flavourful wines[48].

The use of sulphur dioxide (SO_2) also influences yeast function during winemaking. SO_2 is commonly employed as a preservative and antimicrobial agent due to its antioxidant characteristics and ability to inhibit undesirable microbial growth. However, excessive or improper use of SO_2 can have negative effects on yeast activity. Therefore, it is essential to carefully consider and regulate the addition of sulphur dioxide to achieve a balance between its beneficial effects and potential impact on yeast performance[13]. In recent years, there is a tendency to make wines without SO_2, however its properties are not meet by any other additive.

1.7. Deviations

As explained, there are several factors that contribute to the successful progression of the alcoholic fermentation process. However, problems can arise during this process especially in the form of sluggish and/or stuck fermentations. These issues cause significant concerns to oenologists, persisting as common problems in wineries. Stuck and sluggish fermentations may be caused by many sources such as an improper initial YAN, sudden temperature changes, high initial sugar concentration, incorrect sulphur dioxide (SO_2) dose or too acidic initial must[49].

A sluggish fermentation usually is an initial indication of potential fermentation trouble. It may happen in different moments of the process: a slow starting, a normal fermentation becoming sluggish or a whole sluggish fermentation, which usually leads to a stuck fermentation[48]. A stuck fermentation is particularly problematic, as restarting the fermentation process in a winery becomes challenging. This involves the addition of necessary nutrients and/or the recovery of the suitable temperature, and the reinoculation of yeast to reactivate the fermentation. Moreover, as part of yeast stress metabolism, many unwanted and unpleasant metabolites may be released to the wine, which will affect wine quality[50,51]. Additionally, in sluggish or stuck fermentations, with the inoculated yeast having problems carrying out the process, other unwanted microorganism could take advantage and increase their population.

In terms of deviation caused by spoilage microorganisms, possibilities are plenty and frequent in wineries[52]. Various sources contribute to the presence of different yeast genera, including the grape surface microbiota, resident yeast in the winery environment, and the use of active dried yeast (ADY) inoculation. However, imposition of a specific yeast to carry out the alcoholic fermentation may be challenging and needs to be properly controlled. Furthermore, some microorganisms such as LAB may be wanted or unwanted depending on the desired outcome of the wine. Low-acidity wines, for example, are not meant to undergo malolactic fermentation as it would further increase the pH and increase the volatile acidity due to their metabolism leading to an undesired process known as "lactic spoilage"[36]. Nevertheless, the main responsible of contamination in terms of microorganisms are AAB, as their metabolism transforms (oxidises) ethanol into acetic acid, resulting in elevated volatile acidity levels and characterized by a pungent odour at high concentrations[50,53]. Despite being in low concentration in the surface of grapes, when

the skin is damaged their concentration could reach 10^6 CFU · mL^{-1} but the alcoholic fermentation conditions keeps their population under control as AAB are anaerobic bacteria; however, a bad management of wine such as too much aeration during winemaking may promote their growth. Another problematic microorganism is *Brettanomyces*, known for producing unpleasant aromas described as "barnyard, sweat or horsey". Even though it may be present in every step of winemaking, its predominant source and impact in wine is during aging in wood barrels[54].

The presence of unwanted microorganisms not only affects the final quality of the wine due to the release of undesirable metabolites but also creates competition for nutrients needed by the inoculated yeast, leading to sluggish or even stuck fermentations[55].

1.8. Wine analysis

Numerous factors influence the composition of wine, making it a complex beverage. Figure 3 provides a schematic representation of wine composition, illustrating the components and their approximate ratio.

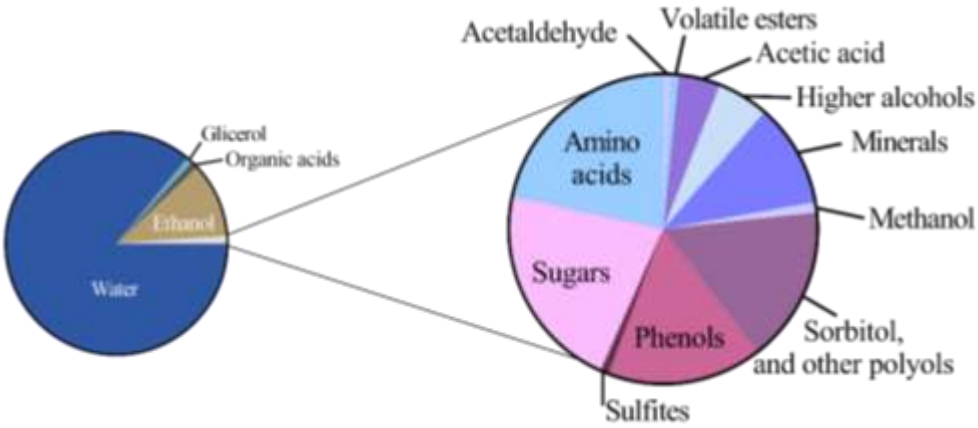


Figure 3. Representative composition of an average dry red table wine (Adapted from Waterhouse *et al.*[34]).

From an analytical perspective, wine is a very complex mixture of many different chemical compounds present in a wide range of concentrations. However, for many winery applications, just specific compounds are analysed based on their importance in specific

metabolic pathways. These compounds or parameters, such as density, pH and temperature, are closely monitored to maintain quality throughout the winemaking process. Many methods have been proposed by the OIV, such as total acidity, volatile acidity, density and ethanol content determination, which are detailed in the "Compendium of International Methods of Analysis of Wines and Must"[56], to facilitate standardized analysis. The OIV compendium serves as a valuable tool, providing validated methods and, many times, alternative approaches in cases where the reference methods may pose difficulties.

Despite the availability of these helpful analytical methodologies, wineries often rely on their small and limited laboratories to perform basic analyses. In some cases, semiquantitative methods are employed solely for comparative purposes. The main parameters routinely analysed in wineries include[57]:

- Sugar concentration: Measured daily, usually through densitometry, to evaluate the progress of alcoholic fermentation. Wineries can assess the process performance by the decrease in density over time.
- pH: Measured using a calibrated pH-meter, it provides a good process performance parameter, as the pH evolves during fermentation due to yeast metabolism. Additionally, pH measurements allow the monitoring of other winemaking processes such as malolactic fermentation and finning, as in these processes, acids are consumed or precipitate.
- SO_2: This antimicrobial, antioxidant, antioxidase compound is widely used in winemaking. Its concentration, either free, combined or total, is carefully controlled both at the beginning of the process (to avoid spoilage by microorganisms) and at the end of the process (to ensure compliance with regulations). Due to its importance in wine legislation, there are many available methods for its determination: Paül method, Ripper method, Toning method, etc.
- Total acidity: At the beginning, to know the starting point of the process and in the end of the winemaking process, it is necessary to control this parameter due to possible low acidity grapes or changes in minor acids. Tartaric acid remains the same through the process as yeasts are not able to use it in their metabolism. The total acid content (excluding dissolved CO_2) is normally assessed with a volumetric

titration to evaluate every compound impacting acidity and expressed as tartaric acid equivalents per litter.

- Volatile acidity: Determining this parameter is crucial both for regulatory compliance and for assessing organoleptic quality. It can indicate potential issues with alcoholic fermentation or microbiological contamination. Specific distillation systems such as the one used by the García Tena method, followed by volumetric titration are commonly employed for this analysis.
- Malic and lactic acids: The evolution of both malic and lactic acids is monitored, particularly in red winemaking, to assess the progress of malolactic fermentation. Thin-layer chromatography is commonly used for the analysis of these acids.
- Organoleptic properties (gustatory, olfactory or visual properties): Traditionally assessed by winemakers, the analysis of organoleptic properties can provide valuable information about various aspects of the wine. When performed by a trained professional, this analysis can offer insights into process performance, contamination issues, and the production of unwanted compounds. Additionally, evaluating organoleptic properties at the end of the winemaking process can help assess the final wine characteristics. To mitigate subjectivity, trained panellists are often employed to ensure more objective evaluations. By considering the wine sensory attributes, winemakers can make decisions to enhance overall quality.

The analytical methodologies used should prioritize speed in obtaining results to minimize the delay between sampling and getting suitable information. This allows timely corrective actions to be taken when necessary. However, there may be complex deviations of the process or specific unwanted organoleptic properties, for which wineries require additional information beyond their in-house capabilities. This limitation implies having to send samples to external laboratories, but this option can be both expensive and time-consuming, which potentially could lead to wrong decision-making that affects overall efficiency of the winemaking process.

2. Spectroscopic techniques

In recent years, there has been a growing interest in the development and application of fast and cost-effective analytical techniques within the food and beverage industries. This growth can be attributed to the emergence of new devices that enable the acquisition of

detailed process information, surpassing the limitations of conventional methodologies[58]. Among the various possibilities, spectroscopy has emerged as a remarkable technique due to its versatile application across multiple fields and its ability to provide a significant amount of information from a sample, often considered as a "fingerprint" technique[59].

Spectroscopy is a set of analytical techniques that involves the measurement, and interpretation of spectra. A spectrum is the result of the interaction between electromagnetic radiation and matter. Electromagnetic radiation arises from the combination of orthogonal oscillations in the electric and magnetic fields. Different types of information about the sample can be obtained depending on the specific ranges of wavelengths of electromagnetic radiation used for the analysis. These ranges divides the radiation in different zones an some of them are very interesting from an analytical point of view such as ultraviolet (UV), visible (Vis), near–infrared (NIR), or mid–infrared (MIR). The interaction between radiation and matter can be of different types including absorption, emission, strokes shift (Raman spectroscopy) or fluorescence. Additionally, the target species for analysis can be molecules to atoms, further expanding the applications of spectroscopy in food and beverage analysis[60].

2.1. Infrared region

The infrared (IR) region of the electromagnetic spectrum covers radiation with longer wavelengths and lower energy compared to the more known visible region. The energy carried by infrared photons induces transitions between vibrational energy states of molecules, leading to vibrational excitations of bonds or specific functional groups within molecules[61]. Analysing the absorption bands in an IR spectrum allows for both quantitative and qualitative molecular analysis by identifying specific chemical groups. While infrared radiation can also cause rotational movements of molecules, these motions are typically superimposed on the vibrational bands and require high-resolution spectrometers for observation[62].

To obtain the infrared spectrum of a sample, a beam of infrared radiation interacts with the sample and the fraction of incident radiation that is absorbed, transmitted, or reflected by the sample at a given energy is determined[63]. This energy is typically represented by wavenumbers or wavelengths, which are proportional or inversely proportional, respectively, to the energy. Only bonds with an asymmetric structure and a dipole moment that changes over time can effectively absorb infrared radiation[64]. The resulting signal

represents the molecular absorption, transmission, or reflection of the sample in the infrared region. Samples with different compositions yield distinct IR spectra, essentially serving as unique fingerprints for each specific sample within the IR region[61].

The infrared spectroscopic region can be subdivided into three regions (Figure 4) depending on its energy power, ordered from less to more energetic: far–infrared (FIR, < 400 cm^{-1}), mid–infrared (MIR, 4000 to 400 cm^{-1}) and near–infrared (NIR, 14000 to 4000 cm^{-1}).

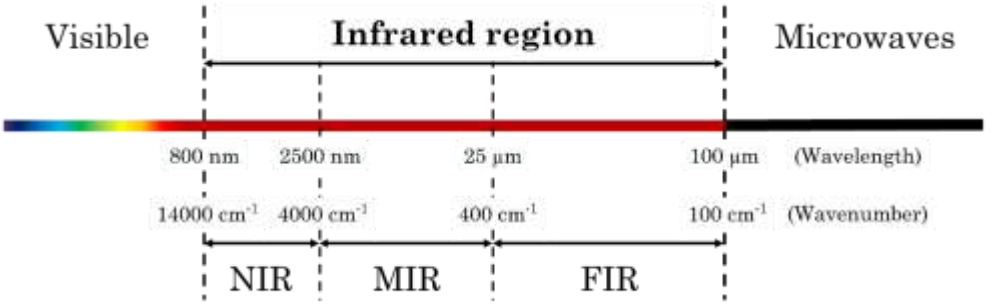


Figure 4. Division of electromagnetic radiation in the infrared region. Adapted from Sun, 2009[61].

Regarding food and beverages, FIR spectroscopy has limited applicability as it only detects vibrations of molecules containing heavy molecules, such as inorganic and organometallic substances[65]. However, NIR and MIR spectroscopies have become highly popular techniques for the analysis of food and beverages[66]. These techniques are related to vibrations of molecular bonds, showed in Figure 5. In MIR spectroscopy the spectra are characterized by showcasing fundamental vibrations with greater absorption capacity and well-defined peaks[66]. On the other hand, the NIR spectrum is the result of combinations and overtones of fundamental vibrations present in molecules that mainly contain hydrogen atoms in covalent bonds with elements such as carbon (C), oxygen (O), or nitrogen (N)[67,68]. These spectroscopic methods provide valuable insights into the chemical composition and molecular structure of food and beverage samples, enabling rapid and non-destructive analysis with potential applications in quality control, authenticity assessment, and process monitoring[69].

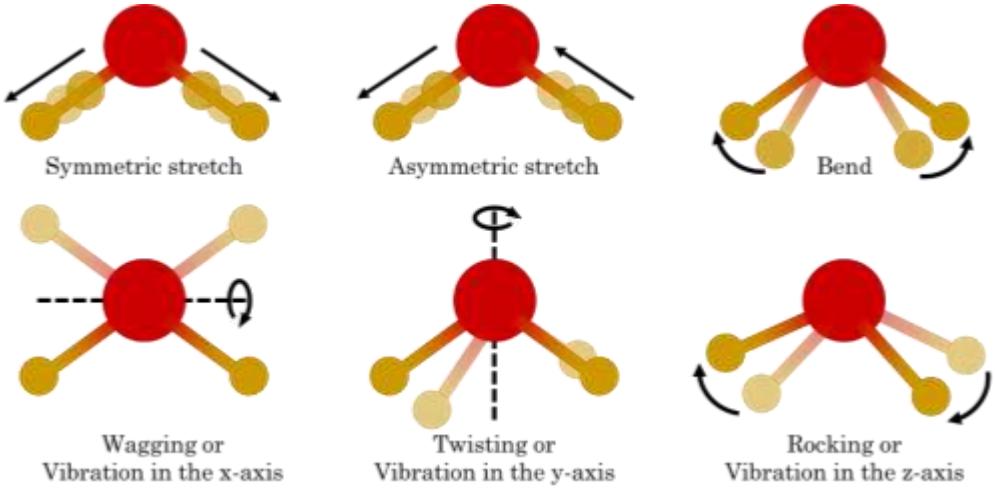


Figure 5. Vibrational modes of the water molecule. Adapted from Rodriguez–Saona *et al.*[70].

2.1.1. Mid–Infrared Spectroscopy

The aforementioned properties of MIR spectroscopy, make it an exceptionally suitable technique for food and beverages analysis. It exhibits high peak resolution, enabling the detection and characterization of molecules even at lower concentrations than NIR spectroscopy. However, the most significant feature of MIR spectroscopy lies in its fingerprint region, spanning from 1500 to 1000 cm^{-1}, which is unique and specific to each sample. The information contained within this region allows for the identification of molecular structures[71].

In MIR spectra, each band can be assigned to a specific functional group, and the intensity of the peak provides insights about the polarity and nature of the peak. Highly polarized covalent bonds exhibit strong absorption in the MIR region compared to less polar or non-polar bonds. Moreover, stronger bonds, determined by their bond strength constant, vibrate at higher frequencies than weaker bonds, as it is necessary more energy to achieve a higher vibrational energy state[63].

MIR spectroscopy faces two significant challenges when applied to samples from the food and beverage industry. Firstly, the high absorption of water can lead to the overlap or hiding of other important peaks. Secondly, the high abundance of molecules presents in

food and beverage samples often share similar bonds, making it difficult to distinguish specific signals for the molecules of interest[61].

2.1.2. Near–Infrared Spectroscopy

NIR spectroscopy, although capable of extracting information from the molecular bonds in food and beverage samples, presents challenges in terms of interpretability compared to MIR spectroscopy, due to the complexity of the samples. This is primarily because NIR spectra are a result of combination or overtone bands of molecular vibrations. In a typical NIR spectrum, broad bands are observed instead of distinct peaks, leading to the spectra of different samples look similar. Moreover, water molecules contribute significantly in the NIR region, manifesting as OH overtones and combination tones at various wavelengths (760, 970, 1450 and 1940 nm)[72].

NIR bands, being derived from overtones and combinations of fundamental vibrations, exhibit weaker signals compared to those obtained in the MIR range. Specific molecular scenarios, such as the CO stretch and NH bonds (which are prominently present in proteins), can contribute to a NIR spectrum with significant band overlap[73]. These limitations had historically restricted the use of NIR spectroscopy. However, technological advancements and computer processing has compensated for its limitations and has facilitated the use of this analytical technique, which offers high cost-effectiveness[74].

2.2. Measurement methods

Sampling techniques have experienced many innovations in terms of optical probes and electronic/computational devices. Nowadays, many challenging problems when measuring any liquid, gas or solid sample have been surpassed, such as water overlapping[72]. IR transmission methods are the most traditional but other measurement configurations have gained popularity, such as attenuated total reflectance (ATR), diffuse reflectance and specular reflectance[75].

Transmission sampling methodology for infrared spectroscopy was the first configuration implemented. Depending on the physical state of the sample different analysis configurations are available. Liquids would need a transmission cell of fixed or variable pathlength and high water absorption requires the use of other solvents for analysis. Gas samples would require filling a glass or brass gas cell; due to small sample compartment

a multireflection gas cell is necessary. Analysis of solid samples have more possibilities depending on the previous sample preparation: alkali halide disc, mulls or films. Due to sample preparation, transmission methodologies are not suitable for *in-situ* analysis of processes[63].

Reflectance methodologies provide a more convenient approach for sample analysis. Depending on how radiation interacts with the sample, there are three different methodologies: specular reflectance, diffuse reflectance, and attenuated total reflectance (ATR)[75].

Specular reflectance is based on the external reflection, where the incident beam and the reflected beam have the same angle of incidence. It is particularly useful for solid samples with flat surfaces and reflective properties, although it can be adapted for use with liquids. Several factors can influence the resulting spectra, such as the angle of the incident beam, surface roughness, refractive index, and absorption properties of the material[76].

2.2.1. Diffuse Reflectance

Diffuse reflectance, also known as Diffuse Reflectance Infrared Fourier Transform (DRIFT) spectroscopy, is also based on the reflectance properties of the sample. However, the incident beam penetrates deeper into the sample than in specular reflectance, and the resulting reflection scatters in virtually every direction, producing diffuse scattered light[70]. This method is well-suited for rough surfaces and powdered samples (often with KBr as a dilutant). The detector collects the beams coming from a wide angle, and the obtained reflectance spectrum correlates the sample concentration with the intensity of the scattered radiation, following the principles of the Kubelka–Munk theory. Figure 6 provides an example of specular reflectance, diffuse reflectance, and a combination of both.

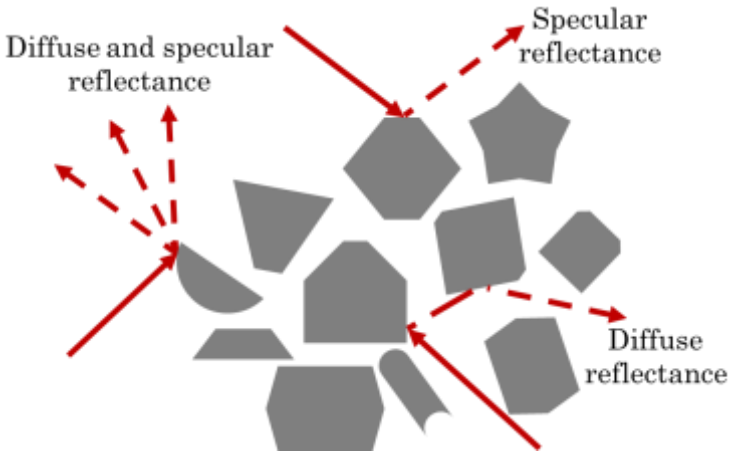


Figure 6. Representation of specular, diffuse and diffuse and specular reflectance in a powdered sample. Adapted from Coates[77].

2.2.2. Attenuated Total Reflectance

ATR is a widely used sampling technique for infrared spectroscopy due to several advantages. One of the main benefits is that it requires minimal or no sample preparation. Additionally, it only requires a small amount of sample, it eliminates the need for cells and avoids variability caused by different path lengths. ATR ensures consistent spectrum collection across measurements for the same sample[66].

The principle behind ATR is based on the attenuation of infrared light when directed at the interface between an internal reflective element (a crystal) with a higher refractive index than the sample. Crystals such as zinc selenide (ZnSe), thallium iodide-thallium bromide (KRS–5), germanium (Ge), silicon (Si), and diamond are commonly used for this purpose. An evanescent wave is generated at the crystal-sample interface, which penetrates the sample. The sample selectively absorbs specific wavelengths, leading to a decrease in the intensity of the reflected radiation. The attenuated radiation leaving the crystal is then measured[63]. An example of ATR for a liquid sample is shown in Figure 7.

Unlike transmission methods, ATR does not require the sample to be thin enough for light transmission. This is because the radiation does not pass through the sample but instead interacts with its surface. The penetration depth of the radiation is limited to a few micrometres (µm), ensuring that the same spectrum is obtained regardless of the amount of sample on the crystal's surface. In the case of liquid samples, intimate contact with the

crystal is easily achieved. For solid samples, pressure is applied to ensure proper contact with the crystal. To address this, pressure clamping mechanisms have been added in spectrometers with ATR, allowing constant pressure on the sample during measurement[70].

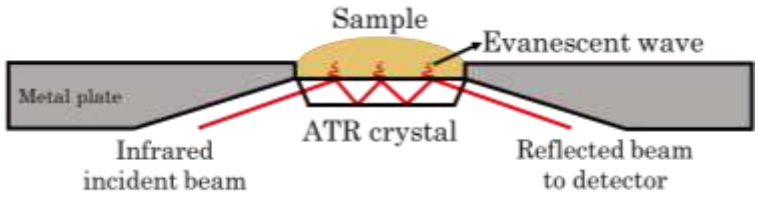


Figure 7. Schematic diagram of the attenuated total reflectance sampling technique.

2.3. Raman

Raman spectroscopy is a complementary technique to IR spectroscopy. In other spectroscopic methods, when the sample absorbs the energy from the incident beam, its molecules reach a temporarily higher energy state before returning to their original energy state. This process is known as Rayleigh scattering. However, a small fraction of the scattered photons (approximately 1 of each 10^7 scattered photons) either goes back to a higher (Strokes shift) or lower (anti-Strokes shift) energy state than the original, being the Strokes shift more usual and the one measured by Raman spectrometers[78]. Such Raman processes occur only for specific vibrations, some of which are exclusive to Raman spectroscopy, while others are observed in both Raman and IR spectroscopy. To obtain a Raman spectrum, a laser with a specific wavelength is used to excite the sample and the resulting Raman scattering is recorded[79].

A typical Raman spectrum is represented by the number of photons per second (y-axis) versus the increasing wavelength (nm) or Raman shift (cm^{-1}). Raman spectroscopy provides a complementary fingerprint to IR spectra. Furthermore, it offers several key advantages. Sample preparation requirements are minimal, or in some cases, unnecessary. Water and ethanol, which are the main molecules in grape must and wine, exhibit weak Raman scattering, making Raman spectroscopy highly suitable for such applications. Additionally, with specific instrument configurations, measurements can be conducted through sample containers. The main limitation of Raman spectroscopy lies in the presence of strong fluorescence in certain samples, which can mask the Raman

scattering signal. To overcome this problem, a longer wavelength excitation laser is employed, as fluorescence is less likely to occur under these conditions[70].

2.4. Spectrometers

IR spectrometers can be classified into two types: dispersive systems and Fourier Transform (FT) instruments. However, FT instruments have largely replaced dispersive ones due to their simpler operation. Dispersive instruments require a radiation source, a monochromator, a sample holder, and a detector (usually connected to an amplifier) to obtain a spectrum. The detector measures the energy at each frequency that passes through the sample, and the resulting signal, representing intensity versus frequency, is plotted[70].

The introduction of FT significantly improved IR spectroscopy by simplifying the spectrometer setup. In FT, the only moving part required is the interferometer[80], so a Michelson interferometer is used, consisting of an infrared light source, a beam splitter (which splits the beam into two), and two mirrors: one fixed and one moving[63]. The split beams are reflected by one of the mirrors and then recombined at the beam splitter. The resulting signal is a result of the interference of the two beams. If the split beams have travelled the same distance, the reflected beams are in phase and all frequencies interfere constructively, producing the highest intensity. However, as the moving mirror changes the distance travelled by the beam, different interference patterns are obtained[71].

After the beam passes through the sample, the obtained signal is called interferogram. The interferogram contains information about the source (located at the centre of the interferogram, known as the Centerburst) and the sample (located in the wings). Initially, the interferogram is not interpretable and needs to undergo mathematical calculations using the Fourier Transform to obtain the desired spectrum. This calculation is performed directly in the equipment and involves breaking down the interferogram into sine waves for each wavelength or wavenumber. FT for IR spectra acquisition has improved the quality of the spectra, reduced the time required to obtain them (as all frequencies are measured simultaneously), and incorporates an internal calibration using lasers as internal wavelength calibration standards. Additionally, its simplicity reduces the possibility of mechanical problems. It should be noted that the acquisition of a background is necessary, which can be air or water depending on the analysis objective and sample

type. The background enables the determination of the reference point for absorption intensity[71].

Furthermore, over the past two decades, technological advancements have made it possible to move measurements from the laboratory to the sample, shifting the paradigm of spectrometry. As a result, the design of spectrometers has evolved, with emphasis placed on achieving a satisfactory fast response for specific applications while considering constraints such as budget and size[81]. In contrast, users, who may have limited knowledge of spectroscopy, primarily seek a user-friendly interface and consistent performance. These factors are not traditionally defined solely by parameters like signal-to-noise ratio or stability over time, but more to robustness to movement, stability over time fluctuations, sealing for particles or liquids, power consumption, heat dissipation and battery life[82].

The portability has been able to be achieved through the apparition and application of some technological advances, such as the diode array instead of the moving mirrors, the use of LEDs as light sources, which also benefits the heat management and the microelectromechanical systems, to achieve smaller instruments. Both improvements have been more notorious and applicable in NIR spectrometers than in MIR ones. Additionally, portable spectrometer prices have diverged regarding to the comparison of NIR and MIR spectroscopy, as the size of the optics and the need of use the moving mirror for MIR, have made that these equipments are still much more expensive and bulkier [82].

3. Chemometrics

The techniques mentioned in the previous section generate a high amount of data per analysed sample, as every wavelength represents a value, accounting for hundreds of values. Additionally, due to their fast response, they allow for obtaining information from a large number of samples in a short time. As a result, high-dimensional data matrices are generated, containing the information about the samples and measurements. For example, when measuring IR spectra for multiple samples, an IxJ matrix is obtained, where I represents the samples and J represents the wavelengths or wavenumbers. Furthermore, if the spectra are taken from a process, such as alcoholic fermentation, a third dimension, K (time), is considered, resulting in an IxJxK matrix. Multivariate data analysis (MVDA) techniques are required to process the data and extract the desired information from the samples, if they have any source of variation or if there are any

aggregation or relationships. MVDA techniques, when applied to analyse the chemical information of samples, are known as "chemometrics"[83].

The first step when dealing with a three-dimensional matrix is data arrangement or unfolding. In the given example, two possible data arrangements can be used. The first one is I*K*xJ, where each spectrum of the samples (I) throughout the entire process (in every time, K) is concatenated ($I \cdot K$) along one matrix axis, while the wavelengths (J) are along the other. The second approach is IxJK, where spectra from the same sample at different times (K) are concatenated ($K \cdot J$) along the wavenumbers (J), while the samples (I) are along the other axis[84,85].

3.1. Data pre–processing

After obtaining a two-dimensional matrix, the next step involves processing the data to enhance their quality by reducing random noise caused by instrument variations, such as fluctuations in light sources and light scattering. This processing aims to minimize sources of spectral systematic variability and improve selectivity, ultimately leading to better accuracy and robustness for subsequent data analysis[86,87]. While each analytical signal has its specific pre-processing techniques, this section will discuss the common pre-processing techniques used in this thesis for IR spectroscopy, categorized as normalization and filtering[87,88].

3.1.1. Normalization

Normalization is a type of data pre-processing that aims to minimize spectral variability caused by variations in light sources or scattering effects[89]. The two most commonly used methods are standard normal variate (SNV)[90] and multiplicative scatter correction (MSC)[91]. SNV involves subtracting the mean spectrum from each individual spectrum and dividing the result by the standard deviation (calculated row-wise). On the other hand, MSC utilizes regression on the mean signal to correct wavelength-dependent effects.

3.1.2. Filtering

Filters are employed to mitigate spectral offsets, positive or negative slopes in spectra, broad baseline distortions, and other baseline effects. Baseline correction encompasses several methods to minimize distortions caused by scattering, which range from subtracting a straight horizontal line to subtracting a polynomial function[88]. Smoothing

methods aim to remove random noise by fitting a polynomial to a range of data points (moving window averaging), and among the smoothing techniques, Savitzky-Golay is the most commonly used[92]. Derivative methods may serve various purposes, such as enhancing small peaks, accentuating shape differences between nearly identical peaks, and correcting baseline shifts and drifts (depending on the derivative order). Savitzky-Golay derivative methods are popular as they apply smoothing and derivatives in a single step[88].

Subsequently, mean centering is commonly applied to conclude this step, as it allows for the correction of systematic differences between variables. Mean centering involves subtracting the mean value of each variable from each raw value, resulting in data with a zero mean. This method preserves spectral shape variations while eliminating differences between large and small peaks along the spectrum[93].

Depending on the objective, other spectroscopic pre-processing techniques have also proved to be useful. However, it is important to assess the pre-processing outcomes by visualizing the changes in spectra (although this evaluation requires prior knowledge of the signal and data) or evaluating model parameters to understand and quantify the effects of the applied pre-processing steps. In some cases, trying pre-processing combinations that have worked in other models can be useful, but this may involve a trial-and-error approach[94].

3.2. Variable selection

When modelling the information, it is important to consider that some variations may be attributed to instrumental noise rather than chemical information from the sample. In order to improve the model, it may be necessary to remove non-transcendental variables, and there are several strategies to accomplish this[95].

Filter methods are applied after model building and involve selecting important variables by ranking the variable and applying a threshold. The Variable Importance in Projection (VIP) and the Selectivity Ratio (SR) are the main methods in this category[95–97].

Wrapper methods involve adding or removing variables and rebuilding the model until no significant difference between the models obtained is achieved. Examples of methods in this category include uninformative variable elimination in Partial Least Squares (UVE–

PLS), interval PLS (iPLS), genetic algorithms, and predictive property-ranked variables (PPRV)[95–97].

Alternatively, variable selection can be directly integrated into the model building process to reduce computation time compared to wrapper methods. Typical methods involve introducing sparsity into weight vectors, regression coefficients, loadings, or canonical vectors[95–97].

In addition to these techniques, spectroscopists may also selectively choose specific regions based on their prior knowledge of the sample and spectra[88,98].

A well-chosen combination of pre-processing and variable selection techniques can greatly improve model performance and interpretability. In model calibration, effective variable selection can lead to a reduction in analysis time, as future samples only need to be analysed in the selected regions.

3.3. Modelling. Unsupervised methods

When performing data analysis, it is usual even mandatory to initially employ unsupervised methods as exploratory tools. Unsupervised methods solely require measured data, the spectra, as input and do not rely on any additional information. These methods are used for various purposes, including data visualization, outlier detection, identification of trends, and grouping based on data variability[99].

3.3.1. Principal Components Analysis

The most commonly used method for data exploration is Principal Component Analysis (PCA). This method focuses on reducing the dimensionality of the variables, resulting in new uncorrelated variables known as Principal Components (PCs). PCs are a linear combination of the original variables (Figures 8 and 9), and they are arranged in order of the highest variability. This means that the first PC explains the direction of highest variability, the second PC explains the second highest direction of variability, and so on. In practice, the interpretation of PCA involves primarily focusing on the first few PCs, as they capture the main directions of variability in the data[100,101].

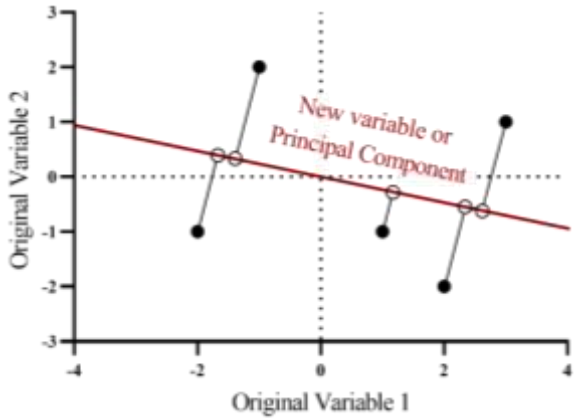


Figure 8. PCA model for two original variables showing the first PC of the model (red line) and the projection of the samples in the PC (circumference). Adapted from Kucheryavskiy[102].

Each PC is defined by the product of two vectors: the score matrix **T**, which represents the projection of the samples onto the new coordinate system, and the loadings matrix **P**, which represents the weight coefficients of the original variables. The information not explained by the new variable is left in the residuals matrix, **E**. Data decomposition in fewer variables can be described by the following equation[103]:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{T} + \mathbf{E} \qquad \text{Equation 1}$$

A schematic representation of the PCA matrix decomposition of the data depicted in Figure 8 is shown in Figure 9.

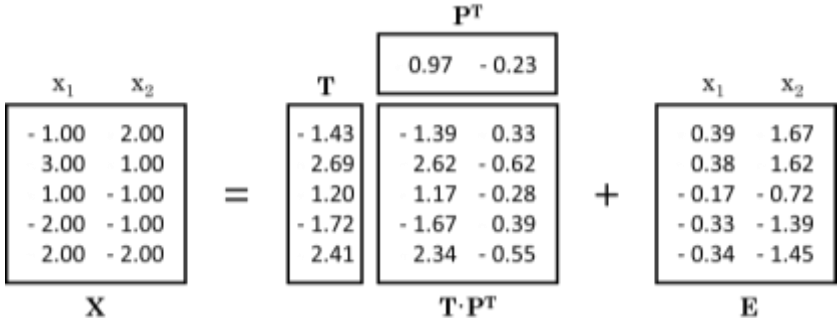


Figure 9. PCA decomposition for the example given in Figure 8 in matrix form. The original matrix (**X**) is decomposed in the product of scores (**T**) and loadings ($\mathbf{P}^{T}$) plus the residual matrix (**E**). Adapted from Kucheryavskiy[102].

3.4. Supervised methods

Supervised methods utilize both measured data and prior knowledge provided by the analyst regarding the sample. This prior knowledge can take the form of empirical values, such as the reference value of a physicochemical parameter, or qualitative information obtained from the analysis, such as the sample group classification or the design of experiment (DoE) employed to generate the samples[99].

3.4.1. Partial Least Squares (PLS) regression

When predicting physicochemical parameters using fingerprint techniques like IR spectroscopy, algorithms need to establish a correlation between the signal (a two dimensional data matrix, normally represented as *X*) and the parameter of interest (a one dimension matrix, normally expressed as **Y**, with the same sample dimension than **X**). The most commonly used algorithm for this purpose is Partial Least Squares (PLS) regression. In PLS, a similar approach to PCA is applied to decompose the original matrix into new variables called Latent Variables (LVs). However, in PLS these LVs maximize the covariance between the signal and the predicted parameter. In the example shown in Figure 8, the direction of variability of the physicochemical parameter (represented by a range of purples based on reference values in Figure 9) may not align with the main direction of variability. Therefore, a compromise needs to be found between the variability of the spectra and the variability of the parameter[104].

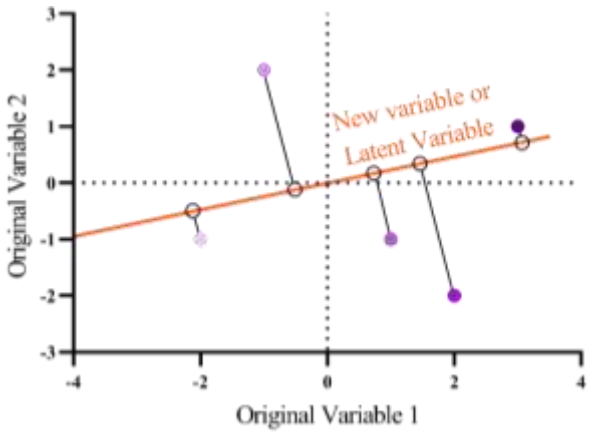


Figure 10. PLS model for two original variables showing the first LV of the model (orange line) and the projection of the samples in the LV (circumference). Different purple shades depict a range in a physicochemical variable for the samples. Adapted from Kucheryavskiy[102].

In the case of the PLS algorithm, **X** and **Y** matrices are decomposed following the equations 2 and 3.

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \qquad \text{Equation 2}$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^T + \mathbf{F} \qquad \text{Equation 3}$$

Where **T** and **U** are the scores of **X** and **Y**, respectively, **P** and **Q** are the loadings related to the variables in **X** and **Y**; and **E** and **F** are the residuals of **X** and **Y** not explained by the selected LVs. Then, the regression vector that relates the information of **X** and **Y** is calculated as:

$$\mathbf{U} = \mathbf{T} \cdot \mathbf{B}_{PLS} + \mathbf{H} \qquad \text{Equation 4}$$

$\mathbf{B}_{PLS}$ is a matrix containing the PLS coefficients to relate **X** and **Y** through their scores. More details on how the regression coefficients are obtained can be found in Wold *et al.*[105]. **H** is the matrix of the information not explained by the model, the residuals matrix.

3.4.2. ANOVA – Simultaneous Component Analysis

A DoE is a data matrix containing the experimental conditions factors specifying the individual categories for each factor. In the case of ANOVA – Simultaneous Component Analysis (ASCA) supervised method, the additional information given with the measured data to the algorithm is a matrix that contains in which category for each factor is considered a specific sample. It can be considered that ASCA is a generalization of ANOVA but for multivariate data. Considering two factors, ASCA decomposes the original (centred) data matrix according to:

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = \mathbf{X}_{Factor\ 1} + \mathbf{X}_{Factor\ 2} + \mathbf{X}_{Interaction} + \mathbf{E} \qquad \text{Equation 5}$$

where **X** is the original data matrix containing the spectra, **1** is a vector of ones, $\mathbf{m}^T$ is the mean of all the observations, $\mathbf{X}_{factor\ 1}$ and $\mathbf{X}_{factor\ 2}$ are the matrices representing the effects of each one of the experimental factors, $\mathbf{X}_{Interaction}$ contains the interaction between the factors and **E** is the residual matrix. Each matrix is centred and contains the mean profiles of the samples corresponding to each factor or interaction level. Since all effect matrices are centred, the magnitude of the effects can be evaluated as the sum of the squared matrix elements. Given a factor *i*:

$$\mathbf{SSQ}_{\text{factor } i} = \|\mathbf{X}_{\text{factor } i}\|^2 \qquad \text{Equation 6}$$

where SSQ is the sum of squares of the elements in the matrix and $\|\cdot\|$ is the Euclidean norm. The interaction matrix is calculated after the subtraction of the main effect matrices. Then, a bilinear decomposition of each effect 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 5 can be decomposed as:

$$\mathbf{X}_{\text{factor } i} = \mathbf{T}_{\text{factor } i} \cdot \mathbf{P}^{\mathrm{T}}_{\text{factor } i} + \mathbf{E}_{\text{factor } i} \qquad \text{Equation 7}$$

where $\mathbf{T}_{\text{factor } i}$ is the score matrix, $\mathbf{P}^{\mathrm{T}}{}_{\text{factor } i}$ is the loading matrix and $\mathbf{E}_{\text{factor } i}$ is the residual matrix of the i^{th} partitioned matrix in equation 5. By reducing the dimensionality of the data, a clearer visualization and interpretation can be achieved, allowing for the examination of each experimental factor or interaction individually. The loadings for factor i define a subspace spanned by $\mathbf{X}_{\text{factor } i}$, which emphasizes the spectral directions associated with the factor being investigated. The scores for factor i are the new coordinates of the observations on the Simultaneous Components (SCs) of the model[106].

Other common supervised algorithms are soft independent modelling of class analogies (SIMCA) for classification[107] and multivariate curve resolution to address mixture analysis problems[108].

3.5. Multivariate statistical parameters

To assess the performance of the model, it is essential to calculate various statistical parameters that provide insights into the sample's classification, predictive ability, and the statistical significance of the results.

Hotelling T^2 and Q residual are particularly useful parameters in determining the sample's similarity to the rest of the samples and its belongingness to the model space[93].

Hotelling T^2, can be seen as an extension of the univariate *t*-test and is applied to the scores obtained from the model. Assuming a normal distribution of scores, it establishes statistical boundaries and quantifies the distance between a sample and the centre of the model[93]. It is calculated as:

$$T_i^2 = \frac{\mathbf{t}_i^T \cdot (\mathbf{T}^T \cdot \mathbf{T})^{-1} \mathbf{t}_i}{I-1}$$ Equation 8

where **T** is the score matrix of the calibration samples, $\mathbf{t}_i$ is a vector containing the scores of the projected *i*th sample in the model space (PCA or PLS model, for instance) and *I* is the dimensionality of the score matrix[93]. The T^2 limit to statistically determine whether a sample has a high Hotelling T^2, meaning an unusual score value compared to the rest of samples, is calculated as:

$$T_{i\,(I,A)}^2 = \frac{A \cdot (I-1)}{I-A} \cdot F_{A,I-A,\alpha}$$ Equation 9

where *A* is the number of principal components, *I* is the dimensionality of the score matrix, and α the significance level. By analysing the Hotelling T^2 value, we can determine how well a sample fits within the model space and identify potential outliers or samples deviating significantly from the model's central tendency[93].

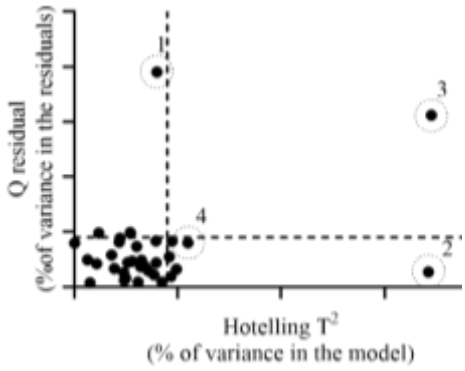
Another important parameter that describes how well a model includes a sample into its space is the Q-statistic, the sum of the squared residuals, that is, the squared sum of the distance between the sample and its projection onto the model space[109]. It is calculated as:

$$Q_i = e_i \cdot e_i^T = x_i \cdot \left(I - P_A \cdot P_A^T\right) \cdot x_i^T$$ Equation 10

where $\mathbf{e}_i$ is the ith row of the residual matrix (**E**), $\mathbf{P}_A$ is the matrix of the A loading vectors considered in the model, and **I** is the identity matrix of the appropriate size. A confidence limit for the Q-statistic can also be calculated as follows:

$$Q_{lim} = \theta_1 \cdot \left[\frac{z_\alpha \cdot \sqrt{2\theta_2 \cdot h_0^2}}{\theta_1} + 1 + \frac{\theta_2 \cdot h_0 \cdot (h_0 - 1)}{\theta_1^2}\right]^{1/h_0}$$ Equation 11

By assessing and evaluating the magnitude of the Q residual, we can identify samples that are poorly explained by the model and potentially require further investigation. Both Q versus T^2 statistics plotted together, is called an influence plot, which is very useful for outlier detection (which could be of different kind as shown in Figure 11) or to detect samples with high influence in the model, especially when there are a lot of samples and normal distribution is correctly assumed[93,100,109].



1 – High residual = the sample is far from the model, but its projection of the model is Near the statistical centre.
2 – High Hotelling T^2 = the sample is far from the centre of the model, but in the same space.
3 – A combination of both, the sample is far from the model space and its projection is far from the model centre.
4 – Sample outside the statistical limit but not considered as outliers.

Figure 11. Influence plot example, showing three case scenarios of different possible outliers (circled with a dashed circumference).

Other important model parameters can be calculated depending on the validation procedure chosen to assess the prediction performance. One such parameter is the root mean squared error of cross-validation or prediction (RMSECV or RMSEP)[104], calculated as:

$$\text{RMSECV} = \sqrt{\frac{\sum_i^{n_{CV}} (y_i - \hat{y}_{i,CV})^2}{n_{CV}}} \text{ or } \text{RMSEP} = \sqrt{\frac{\sum_i^{n} (y_i - \hat{y}_i)^2}{n}} \qquad \text{Equation 12}$$

where y_i is the measured value, $\hat{y}_i$ or $\hat{y}_{i,CV}$ is the predicted value for the internal (RMSECV) or external (RMSEP) validation respectively and n_{cv} or n is the number of total samples used in the total number of cross-validation samples or the external validations. Internal or external validation are two ways of validating the prediction model, assessing the performance of the model with samples not used in the calibration. The difference between both strategies is that in cross-validation or internal validation a subset of the samples is taken out, then the model is build, the excluded samples are projected, and the performance of the model is assessed; then, the samples are included back and another subset is selected to repeat the process[104,110]. For external validation, a subset of samples is excluded from calibration, and used just to project into the model and assess its performance. Samples are not considered in this case back into calibration.

Other parameters are often needed to contextualize the prediction errors, such as the Ratio of Performance to Deviation (RPD), which considers the ratio between the standard deviation of the predictand and the Standard Error of Prediction (SEP). This parameter

helps account for the increase in RMSEP due to an expanded measurement range, thereby removing range effects[111]. It is calculated as:

$$\mathrm{RPD} = \frac{\mathrm{SD_y}}{\mathrm{SEP}} \qquad \text{Equation 13}$$

Several authors have proposed different categories to define model performance, with an RPD above 2 often considered a threshold for good model performance. Throughout this thesis, this value has also been used to assess model performance. However, caution should be taken when using this parameter, as biased or non-normally distributed models may still exhibit good RPD values[112].

Another parameter used for range standardization of RMSEP values is the Range Error Ratio (RER), which takes into account the maximum (y_{max}) and minimum (y_{min}) values[111], and is calculated as follows:

$$\mathrm{RER} = \frac{y_{max} - y_{min}}{\mathrm{SEP}} \qquad \text{Equation 14}$$

Throughout the thesis, a RER higher than 10 has been considered indicative of good model performance. It is important to note that this parameter should also consider the assumption of normal distribution and the statistical significance of extreme samples within the overall sample set[112].

Finally, when working with DoE in ASCA, a permutation test is performed to assess whether the observed impact of a designated factor is a result of the factor itself and not a random result. This test involves comparing the distribution of results after permuting the DoE under the null hypothesis. The permutation test allows the calculation of the significance of a factor without the constraint of normality of samples values[113]. It is calculated as:

$$p - \mathrm{value}(\mathbf{X}_i) = \frac{\mathrm{nbr}\left(\mathbf{SSQ}(\mathbf{X}_{i,\mathrm{perm}}) \geq \mathbf{SSQ}(\mathbf{X}_{\mathrm{factor}\,i})\right) + 1}{k+1} \qquad \text{Equation 15}$$

where "nbr" is the number of occurrences, k is the number of permutations and $\mathbf{X}_{i,perm}$ is the matrix obtained after a random row permutation. The *p*-value obtained from the permutation test indicates the number of cases where the variance of a specific factor is lower than the variances resulting from permutation[106,114]. In the thesis, 10,000

permutations have been used for each ASCA model, which sets the limit for statistical significance. A *p*-value less than 0.05 indicates that only 500 permutations have resulted in better results than the actual factor distribution.

3.6. Multivariate Statistical Process Control

Process control is an essential requirement in every industry to ensure proper process behaviour and to understand the reasons behind process faults. When statistical methods are used for monitoring, it is known as statistical process control (SPC). By closely monitoring the process, the quality of the final product can be ensured, and production can be improved. Identifying the cause of process deviations enables early corrective measures to be taken, thus minimizing the impact of these deviations on the final product's properties and quality[115].

Industries implement SPC using control charts, which graphically represent process performance over time. These charts, aided by statistical limits, help in understanding and predicting process behaviour, identifying possible deviations, and characterizing different types of variation.

As spectroscopy arose, fingerprint techniques have been used that gather more information about the process than univariate techniques. As more information is obtained new approaches or the adaptation of existing approaches are required. This techniques that deal with multivariate information of a process to monitor and control it are known as multivariate SPC (MSPC). In MSPC the first step is data reduction since creating control charts for hundreds of variables is tedious, impractical and ineffective[116]. With MSPC, the new variables obtained through linear combinations of the original ones provide a more comprehensive information, including interactions and combined effects, in a single control chart[117].

Several approaches can be employed to construct MSPC control charts. The most common and straightforward method involves using PCA to reduce dimensionality and analyse the scores, Hotelling T^2 or Q residual values or their evolution over time. Processes that have not suffered any problem or deviation, known to work under normal operating conditions (NOC) are used to construct these charts. Control charts based on various statistical values and limits can be used individually or together, as they complement each other. Statistical limits can be calculated using stablished formulas, such as for Hotelling T^2 and

Q residuals, or calculating the confidence interval or the value of the confidence level times the standard deviation.

For future processes to be monitored, if the behaviour remains under control, their projections on the control charts should fall within the limits. However, if a process deviation occurs, the chart will show an evolution crossing the statistical limits, indicating a deviation from normal behaviour. Additionally, early signs of process deviations, even before crossing the limits, can be detected, triggering early intervention measures. Studying the contribution of the measured signal (using a fingerprint technique) allows determining the cause of the deviation and acting consequently to dismiss the fault[118].

MSPC control charts can also be created by predicting a specific parameter, particularly when industries must comply with regulations regarding maximum allowable concentrations in a process. In such cases, prediction models using methods like PLS are employed. A comprehensive analysis of the errors across the parameter range is conducted to establish reliable statistical limits that ensure compliance with the regulations, considering the multivariate nature of the errors. These prediction-based control charts provide valuable insights into the process performance and enable proactive measures to maintain compliance with the legal requirements[119].

4. Wine fermentation monitoring

The application of spectroscopy in monitoring wine alcoholic fermentation is established on the theoretical basis discussed in previous sections. Figure 12 shows three examples of using different spectroscopic techniques, namely MIR, NIR, and Raman spectroscopy, to observe the spectral differences in grape must and wine. These observed differences in the spectra, which would occur from the beginning to the end of the alcoholic fermentation process, offer potential opportunities when employing spectroscopy as a tool for process monitoring. Moreover, it opens up the possibility of developing MSPC tools specifically tailored for monitoring and controlling the fermentation process, further enhancing its efficiency and final product quality.

The initial stage in implementing MSPC involves gaining a thorough understanding of the process at hand. To achieve this, numerous researchers have investigated into various aspects of the winemaking process, particularly focusing on predicting different parameters at different stages. Given that vinifications are typically conducted on a large

scale and are seasonal processes, it becomes crucial to maximize the number of replicates, explore various conditions, and even simulate process deviations. Consequently, initial investigations of the process are often carried out on a laboratory scale under controlled conditions. Setting up such studies can present challenges in adequately capturing the complexities of the process. Since literature focuses more on the explanation of the results obtained than on the preliminary trials before doing an actual experiment, the following tutorial was developed to provide researchers with a starting point for the study of alcoholic fermentation, specifically employing portable MIR spectroscopy.

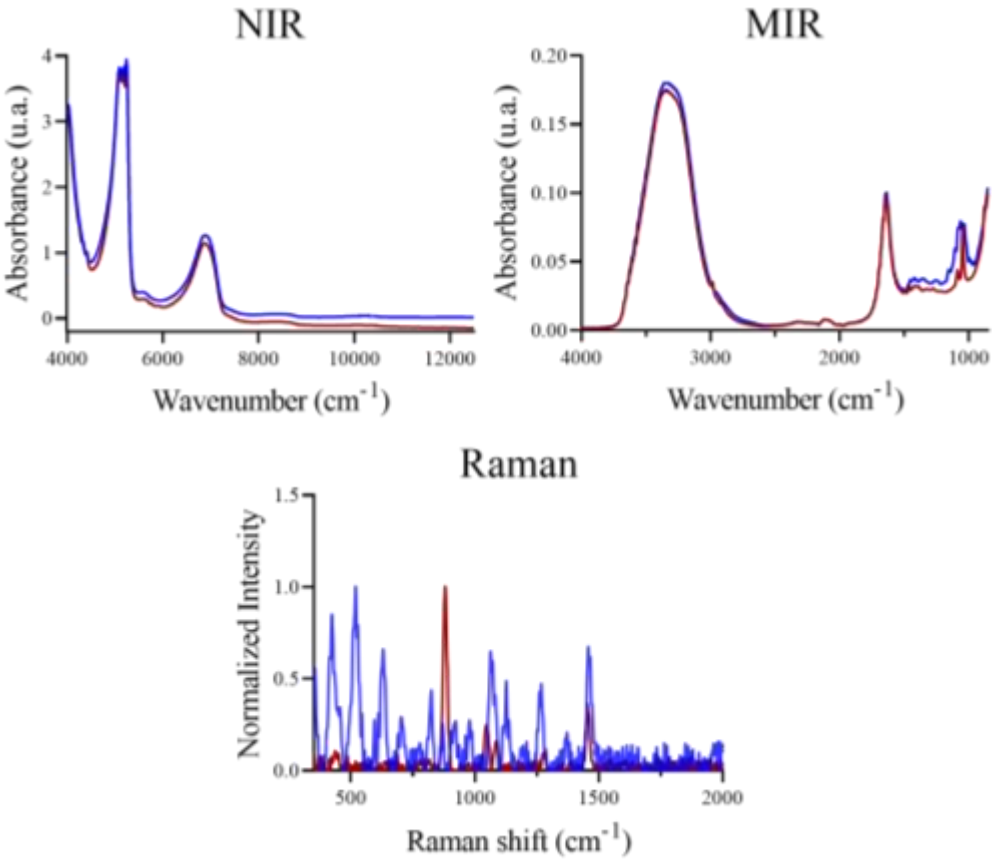


Figure 12. Spectra of grape must (blue) and wine (red) in NIR, MIR and Raman spectroscopy.

ATR–MIR spectroscopy as a process analytical technology in wine alcoholic fermentation – A tutorial

Daniel Schorn–García[a], Julieta Cavaglia[a], Barbara Giussani[b], 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, 43007, Spain
[b]Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, Como, 22100, Italy
[c]Universitat Rovira i Virgili. Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Tarragona, 43007, Spain

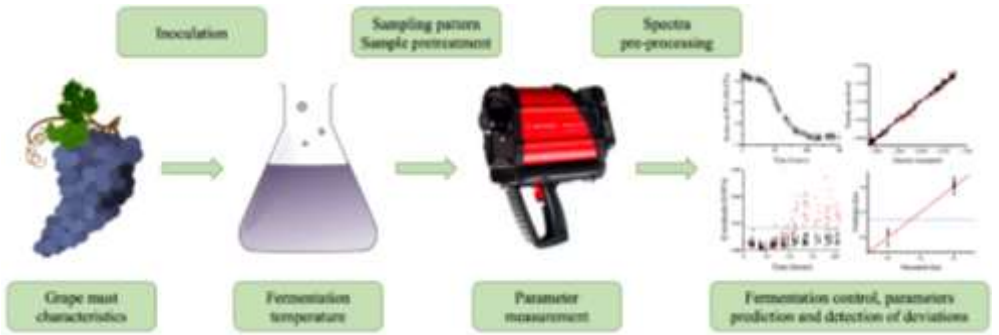
Microchemical Journal 166, 106215 (2021)

Abstract

The goal of this article is to guide the reader through the critical points to be faced when monitoring a fermentation following a Process Analytical Technology (PAT) approach. To achieve this purpose Attenuated Total Reflectance – Mid–Infrared (ATR–MIR) spectroscopy coupled to chemometric techniques are proposed.

Each of the crucial steps (set up of microvinifications, sampling, spectroscopic analysis and chemometric data treatment) is deeply investigated, revealing how the sampling is decisive for the subsequent modelling phase, suggesting how to set parameters to obtain good quality signals, and explaining how to prepare the data for the chemometric modelling and to perform the calculations. The modelling strategies here presented, based mainly on basic chemometric tools such as principal component analysis and partial least square regression, proved to be effective to the purposes and affordable even for non-expert chemometric users.

The article shows, using real examples, how to obtain or predict several parameters from a fermentation data set - control of the fermentation evolution, prediction of oenological parameters during the alcoholic fermentation and detection of deviations from the normal operation condition.



Keywords

infrared spectroscopy; PAT; microfermentation; wine; multivariate process control

Introduction

Wine production is based on the biochemical process called alcoholic fermentation, which implies the transformation of sugars (glucose and fructose) into ethanol and carbon dioxide carried out by yeast, usually of the genus *Saccharomyces*. It is a complex process due the release of many yeasts' metabolites at very different concentrations and that are related to physicochemical and organoleptic properties of the final product[1]. Thus, the control of these by-products has a substantial impact on the final quality of the wine. However, apart from the daily visual observation of the tanks, the parameters that are routinely and traditionally measured in the cellar to follow the alcoholic fermentation are simply the temperature, the density and the pH. Although there are already some simple devices on the market that can measure and predict these oenological parameters, they are not widely used. This is because, as in determining these parameters using other analytical methods, there is a time lag between obtaining the results and applying the corrective actions. Thus, when problems such as deviations, stuck or sluggish fermentations are detected, the delay in carrying out the corrective measures negatively affects the quality, which in the end means economic losses[2]. This is why there is a growing interest in an effective monitoring strategy based on a real-time approach[3].

The United States Food and Drug Administration defined Process Analytical Technologies (PAT) as "a mechanism to design, analyse, and control manufacturing pharmaceutical processes through the measurement of Critical Process Parameters which affect Critical Quality Attributes" [4]. This philosophy is well suited to any other industry where quality control is essential, such as the food industry. It should be noted that food samples are often complex chemical mixtures and, many times, easy to be altered. Therefore, the quality of a food product cannot be guaranteed by analysing only the final product, but a control of the product is required throughout the production process by in-line or on-line measurements. Due to its feasibility and practicality, the PAT approach has already been applied to different sectors of the food industry with satisfactory results (see e.g. Jerome *et al.*[5], van den Berg *et al.*[6] and Pu *et al.*[7] and references therein). As can be seen, in all cases, in the Food PAT approach it is imperative to fully understand the process and understand all the sources of variability to obtain reliable results. Concerning wine alcoholic fermentation, sources of variability come from both raw materials (grape variety, sanitary state of the grape, ripening and vintage) and the fermentation process (species

and strain of microorganism, sudden temperature changes, excess or absence of oxygenation, amount of assimilable nitrogen or spoilages, among others).

As the sources of variation in biochemical processes can be many and very different, several parameters should be determined to detect the potential problems and, usually, these parameters should be determined in the different batches at several production points. Fortunately, the use of modern multi-analysers greatly simplifies these determinations, although the large amount of data generated is usually difficult to manage. So, to extract valuable information these data should be processed through proper statistical/chemometric tools[8]. This is why the PAT approach often relies on the use of multivariate analysis in order to deal with the maximum amount of data to better control the process. Multivariate analysis allows both qualitative (classification or pattern recognition) and quantitative results (parameter prediction) through the analysis of the generated data[3].

Among the different instrumental techniques used for food process monitoring, the spectroscopy-based analysers are taking on a special relevance. The rapidness of the data acquisition, the absence or minimal sample pre-treatment and the portability of the instruments make spectroscopy an ideal tool in PAT approaches. This last characteristic is especially interesting because it allows moving the equipment to the sample instead of having to take the sample to the laboratory. In other words, with this type of equipment we can get the laboratory to the measurement point, thus saving time, improving the efficiency of the testing process, and allowing a decision to be made on the spot.

Although NIR spectrometers have been widely used in food process control[9,10], ATR–MIR (Attenuated Total Reflectance – Mid Infrared) spectroscopy has gained popularity in food analysis as it is fast, non-destructive and environmentally friendly[11] and allows to surpass the water absorption limitations that other types of IR measurements present.

The aim of this tutorial is to help the reader understand the alcoholic fermentation of wine from the PAT point of view, and to provide a guidance on how this process can be controlled using ATR–MIR spectroscopy and what information can be obtained from such control. Following the PAT guidelines and in order to acquire knowledge about the process, this tutorial describes the fermentation process from the beginning and through microvinifications[4]. Thus, the sampling strategy, the analytical procedure, and the multivariate data pre-treatment and modelling will be considered, explained and

discussed and pros and drawbacks will be highlighted for each step. Real examples will be provided to illustrate the main ideas.

Experimental setting and sampling

On-line or in-line[8] measurements are the final goal in process monitoring, but often the difficult direct implementation of instrumentation in the cellar when dealing with high volumes, the need for control experiments and, sometimes, the lack of knowledge about key points in the process discourage the attempt. In this context, micro-scale analysis could be a suitable option for this purpose[4]. In addition, it should be borne in mind that the wine production in a cellar involves a limited number of fermentations, with a subsequent reduced number of batches that will be used to optimize the chemometric strategy. Instead, micro-scale fermentations generate a greater number of available samples to study the process variability, to choose the best strategy to analyse and control the process evolution and to finally facilitate the scale-up of the monitoring process in the cellar. This is even more important when studying complex bio-transformations such as alcoholic fermentation, as biological replicates in different containers have slightly different behaviours.

Grape must characteristics

The importance of the must quality is obvious and, therefore, the legislation allows wineries to adjust some basic parameters of this raw material such as acidity, sugar content or assimilable nitrogen content[12]. When this process is transferred to micro-scale fermentations, they can be produced with natural grape juice or concentrated grape must (CGM). The latter, with a sugar concentration of about 800 $g \cdot L^{-1}$, allows better preservation of the samples before dilution, as it reduces the activity of microorganisms due to osmotic pressure[13]. Moreover, whereas natural grape juice is a seasonal product, CGM can be purchased at any time of the year. The CGM must be properly diluted to reach a sugar concentration that provides the desired alcoholic strength (in a winery the usual concentration is between 170 and 200 $g \cdot L^{-1}$)[1]. Apart from sugar, the pH value should also be adjusted with some acid allowed by legislation, such as tartaric acid, as the pH range of the wine is set between 2.8 and 4.2[13,14]. Finally, it is also necessary to control the Yeast Assimilable Nitrogen (YAN), as it is a limiting factor for yeast growth. Its concentration may range from 140 to 500 $mg \cdot L^{-1}$, with an optimal interval from 200 to 350 $mg \cdot L^{-1}$. A deficit in YAN concentration leads to unfinished fermentations or slower fermentation

kinetics; and a higher concentration would lead to the generation of unpleasant aromas[15]. The addition can be made using ammonium salts or commercial YAN supplements, as it is done in wineries, which apart from inorganic and organic nitrogen content, contain other compounds such as vitamins that enhance yeast fermentation[15].

Inoculation

Once the diluted CGM or natural must have the right composition, the next step is to inoculate the yeast. This can be performed by adding liquid culture media or dry active yeasts to each fermenting container, the latter option being more complicated because it involves an additional step for rehydration. In any case, the temperature difference between the must and the liquid inoculum must be carefully controlled because the suppliers recommend a difference of about 5 ºC and never exceed 10 ºC difference. As for the inoculum concentration, this value depends on the desired fermentation process. Thus, to reproduce the natural population in the grapes to give rise to a spontaneous fermentation it should range between 10^3 and 10^5 CFU · mL^{-1} and to reproduce the inoculum population used in wineries to ensure the imposition of the yeast of interest it should be between 10^6 and 10^7 CFU · mL^{-1}. It should be noted that, although the main yeast species used to inoculate in wineries is *Saccharomyces* due to its high fermentation capacity, it is known that non-*Saccharomyces* yeasts can also influence the course of fermentation and the character of the resulting wine. Recent studies have evaluated the use of controlled mixtures of both species[1,16].

Fermentation temperature

Among the physical parameters, temperature is possibly the most important because the alcoholic fermentation is an exothermic process and yeasts only have proper metabolic activity within optimal temperature ranges. Thus, this parameter deeply influences the kinetics of fermentation, so that when the temperature rises 10 ºC, the speed of the kinetics doubles[13]. Therefore, the fermentations must take place at a controlled temperature, being the recommended values between 10 ºC and 32 ºC. These values differ when it comes to white or red must: in white wine, the temperature should be kept below 18 ºC for aroma retention, but for red must the temperature should be above 25 ºC to improve the colour extraction from pomace[13]. A slight increase in kinetics can also be obtained by agitation; however, to reproduce real conditions in a winery, it is not necessary.

Sampling times

Sampling time is crucial in a process control approach and there are several points to consider:

– alcoholic fermentation follows a sigmoid function, due to the yeast adaptation time at the beginning and the stressful conditions at the end[13,17]. Therefore, to properly define the fermentation curve, a representative sampling at all different phases of the fermentation process is necessary;

– sampling times must be frequent enough to allow corrective actions to be taken in case of deviation from the normality of the process and, from a practical oenological point of view, this usually means less than 12 hours between samplings;

– sampling must reproduce actual conditions, so it should be noted that when using at-line analysis, sampling requires the physical removal of a fraction of the sample to analyse it. This issue is not really crucial for spectroscopic techniques, as the analysis is very fast and requires a small fraction of the sample;

– when applying chemometrics, the sampling design will condition future samplings, as the same number of samplings or even the same sampling frequency will be required to take decisions using the statistical model. If possible, samples should be taken at a fixed frequency throughout the fermentation, but this is not obvious. An effective way to overcome this problem, especially when historical data are used, will be shown in section 5.4.

Sample pre–treatments

One of the main advantages of using vibrational spectroscopy is that little or no sample pre-treatment is required. However, as sampling is carried out from the fermenting must with many microorganisms, to avoid biochemical changes in the samples, it is necessary to remove the microorganisms. This removal can be achieved using physical or chemical pre-treatments. In the case of physical pre-treatments, it is possible to distinguish between partial removal by centrifuging the sample and separating the supernatant, or total removal by filtering the sample through a 0,45 μm (yeast sterilization) or 0,22 μm (yeast and bacteria sterilization) diameter filter. The chemical pre-treatment usually applied is the addition of sodium fluoride (NaF) as an oxidative metabolism suppressor for yeast[17].

For right-after infrared and standard physicochemical analysis, partial removal is a good option (see e.g. Buratti *et al.*[18] and references therein).

Measurements and data collection

The infrared spectral region is characterized by the vibration of the molecular bonds. In the case of the mid-infrared region, approximately between 4000 and 400 cm^{-1}, the spectrum obtained results from fundamental vibrations, which facilitates its interpretation compared to the NIR spectrum. However, it has to be pointed out that MIR spectroscopy shows its full usefulness when the ATR sampling technique is applied. This is because ATR allows to overcome the limitations that imply the large absorptions of water bands[19]. In this tutorial we suggest the use of an ATR–FT–MIR portable spectrometer to monitor the wine fermentation process. This equipment offers the possibility of measuring in real-time and on-site, which also means a reduction in the costs of materials, transport, and storage of samples. In addition, portable spectroscopic instruments are cheaper, have lower operating costs and require less power consumption than their benchtop-equivalents[20]. However, there are also some drawbacks to consider when using a portable ATR–FT–MIR spectrometer. In fact, they tend to record noisier spectra than the benchtop ones and are sometimes considered less accurate and less reliable. Nevertheless, depending on the samples and on the problem to be addressed, this higher noise does not necessarily imply poorer performances when data are properly treated with chemometric tools [see e.g. Crocombe[21]].

From all these considerations, it follows that there are different parameters that condition the spectra obtained. This is why in this tutorial we propose the use of an ATR–FT–MIR portable spectrometer (4100 ExoScan FTIR, Agilent, California, USA) focusing the attention on the parameters to be controlled and optimized to register good quality spectra.

Regarding the usual values of spectral resolution and number of acquisitions, these range between 2 and 32 cm^{-1} and from 32 to 512 scans, respectively. In general terms, the lower the first parameter and the higher the second, the better the signal quality in terms of signal-to-noise ratio. Nevertheless, the time of analysis must also be taken into account, as it increases when increasing the number of scans to get the best signal, so a compromise is needed between the two[22]. In addition, when a 2 cm^{-1} resolution is used, the spectra obtained often show a larger baseline noise that requires a strong smoothing treatment before data analysis with the subsequent loss of information.

The MIR region between 1800 and 900 cm^{-1} has been reported to be the fingerprint of the alcoholic fermentation signals, as the main biochemical bonds absorb in this area[23], so many studies have focused on this region to build the models[18,24,25]. However, there are other important bands in the spectrum related to water and ethanol that can also be used to build the models, as shown by Cavaglia *et al.* when, to predict the pH, the whole spectral range was needed to obtain good results[18].

Other important points to be considered are the cleaning of the ATR crystal and the background collection. In addition to good equipment maintenance, cleaning must be done after each measurement to avoid carry-over effects between samples[6]. It is crucial especially at the beginning of the alcoholic fermentation, when the sugar concentration values are at their maximum, making the sample more viscous. However, the great solubility of sugars, acids and alcohols in water makes cleaning with deionized water quite successful. It is important to evaluate cleanliness based on the results obtained with the background spectra recorded after each cleaning step.

Concerning background collection, two strategies have been described in the literature. In the first one, the background is collected with the crystal empty, providing what is known as an air background[26]. In the second option the background is collected by using deionized water so that the signal obtained allows to subtract the water contribution from the sample spectrum[27]. To our experience an air background has proven to be effective[22].

Finally, it is important to remark that, as IR lamps heats up over time, an equilibrium time is necessary. To be more precise, when switching on the equipment, the spectral intensity changes slowly in such a way that, in order to obtain a stable signal, in the case of our portable ExoScan FTIR instrument we have to wait approx. 100 minutes, as can be seen in Figure S1[12]. Stabilization time is highly dependent on the FTIR equipment and can be lower for equipment with temperature control.

Data Analysis

Spectra pre–processing

The raw signal must be pre-processed to make it more suitable for multivariate analysis and to obtain good quality chemometrics models. It has been shown that the application of optimal pre-processing is critical for PAT, as the resulting models can increase their

performance by up to 25%[28]. In the case of FTIR spectroscopic data, the main objective of pre-processing is to ensure that the signal follows the Beer's law as much as possible[11]. This means removing or minimizing all sources of variation not related to the process studied and this includes removing noise, offsets and baseline drifts and light scattering effects. Typical pre-processing methods used for MIR spectra are divided into two groups: 1) scatter correction methods, which include Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC) and baseline correction; and 2) derivation methods, which include Savitzky-Golay smoothing/derivative as the most used[26,28]. SNV is often used in ATR–MIR spectra to remove the variability due to physical aspects or equipment characteristics, which may generate scatter. Savitzky-Golay smoothing coupled with derivative is useful to reduce baseline noise and simultaneously emphasize small peaks when needed[26]. In our case it is advantageous because the spectra show high absorbance peaks due to water, sugar and ethanol depending on the fermentation time, while peaks related to compounds indicating an abnormal fermentation (e.g. lactic, acetic or malic acids) are smaller and quite overlapped. The smoothing window must be carefully evaluated because a too severe smoothing can remove useful information contained in the spectra.

Model building

Once the data are pre-processed, multivariate data analysis is applied to extract the information from the data, in this case from the spectra. Data analysis aims at finding relationships between samples and variables, detecting trends in the data, and also detecting anomalous samples[29]. The different multivariate analysis methods available can be classified, depending on their purpose, into exploratory, classification or prediction methods, among others.

Exploratory analysis is usually performed as a first step prior to more complex data analysis to provide some preliminary information, such as the presence of sample groupings or the existence of outliers, or to evaluate the precision (repeatability) of the measurements. The gold-standard exploratory technique is principal component analysis (PCA). PCA builds a set of reduced principal components (PCs), which are linear combinations of the original spectra (**X** matrix) and keep the maximum information contained in them. This way, simple two-dimensional plots (scores and loadings) can be plotted to visualize the data.

Classification techniques aim at finding a criterion to assign a sample to a predefined class (or category), each class representing a group of samples sharing specific characteristics[30]. Some of the most used classification techniques are Linear Discrimination Analysis (LDA), Partial Least Squares – discriminant analysis (PLS–DA) and k nearest neighbours (kNN).

Finally, quantitative predictive modelling aims at correlating a set of spectra (**X** matrix) to one or some properties (**Y** matrix). A training set of samples is used to build and optimise the model (where **X** and **Y** are known) and the final objective is to predict the Y properties of a new and unknown set of samples[6]. The multivariate predictive technique most commonly used is Partial Least Squares Regression (PLSR). Real examples using the three model categories are shown from Section 5.2 to 5.4.

Model validation

In all methods, a critical step is the selection of the optimal number of latent variables (LVs). In the case of PCA, the simplest way is to choose those that explain a given percentage of the variance in the spectra (i.e. 95%), although in some cases some LVs accounting for the remaining variance may be also informative. Cross-validation strategies can also be applied (see below). In the case of classification models, and in particular of PLS–DA models, the choice of the optimal number of LVs is performed based on the maximum number of correctly classified samples. i.e. accuracy, in the validation set. However, depending on the problem at hand, sometimes a balance between sensitivity (proportion of out-of-control samples correctly classified) and specificity (proportion of under control samples correctly classified) is sought[31]. Finally, for prediction models, the optimal number of latent variables is selected based on the minimum value of prediction error, estimated for the validation set.

The selection of the optimal number of latent variables is performed during method validation. There are basically two approaches for method validation, depending on the number of samples available: cross-validation and test set validation. Cross-validation is usually applied when the number of samples is low. In cross-validation the original dataset is split in different blocks. Then each block is left-out once at a time, the model is built with the rest of the samples and the left-out block is predicted. This process is repeated for each block and for different LVs. Test-set validation consists on leaving out a subset (usually 30–40%) of the original data, build the model with training set and decide the optimal number of LVs based on the prediction of the test set. The test samples must fall

within the calibration range and have similar physicochemical characteristics. Finally, the number of misclassified samples (for PLS–DA) or the residuals, that is the difference between predicted and measured values (for PLSR) are calculated and a global estimator of the validation error is obtained. For classification models, these estimators are accuracy, sensitivity and specificity, and for prediction models we have the Root Mean Square Error of Cross-Validation (RMSECV) or the Root Mean Square Error of Prediction (RMSEP), the latter for test set validation.

Cross-validation is an important part of model building to avoid over-optimistic models[32]. There are different types of cross-validation, which are distinguished from each other depending on the pattern used when selecting the samples to validate the model. As examples, leave-one-out cross-validation is mainly used for small data matrices, as just one sample per time is used to test. For time-organized data, venetian blinds are useful to assess non-temporarily errors while contiguous block asses temporal stability. Finally, typical errors in CV setup include making only one split of the data, remove small groups of samples or splitting unnatural replicates into calibration and test groups.

Variable selection

Nowadays analysers provide hundreds of data for every sample in a very short time, but sometimes there may be spectral regions not suitable for modelling, that is, not related to the problem at hand. Variable selection (VS) is then an important part in method validation, as it allows detecting the specific spectral regions of interest and produce better models. VS can be performed using algorithms or using chemical knowledge. In the latter case, the operator knows that a certain region of the spectra could be attributed to a given family of molecules. In many cases, a combination of both strategies is applied to obtain better results. There are many algorithms available for VS, such as genetic algorithms, interval PLS, recursive PLS, selectivity ratio or variable importance in projection (VIP), to name just a few (see Westad *et al.*[33] and references therein). All algorithms seek to obtain the best combination of variables, and at this point validation is fundamental (as explained in the Model Validation section) to avoid overfitting (the model just explains the actual samples and no future samples).

Alcoholic fermentation monitoring. Real examples

Data arrangement

The data collected from a fermentation process consist of a series of spectra measured at certain time points and can be structured in a two-dimensional matrix, **X** (*K*x*J*), where *K* are the sampling times and *J* the wavenumbers, which are usually expressed as hours from the inoculation time (time 0).

If data are available for several samples, then data are structured in a three-dimensional matrix, **X** (*I*x*J*x*K*), where *I* samples. This three-way matrix includes the process variability and from it, a NOC (Normal Operation Conditions) space can be defined. It is worthwhile to note that the three-way matrix can be generated only if the sampling times are the same, that is, when spectra are collected at the same time after the start of fermentation, or at the same time intervals.

Depending on the goal of the investigation, the three-way matrix **X** (*I*x*J*x*K*) is usually re-arranged (unfolded) either time-wise (*IK*x*J*) or batch-wise (*I*x*JK*)[9]. In the following sections we will discuss what can be obtained when monitoring a fermentation process by ATR–FT–MIR spectroscopy and the different unfolding and some of the most used modelling strategies necessary to achieve the goals.

5.2. Alcoholic fermentation monitoring

The evolution of the fermentation can be visualized in an easy-to-see way using unfolded PCA (Principal Component Analysis) modelling. PCA compresses the information contained in the original variables (i.e. spectra) into principal components (PCs) to better visualize, usually through bidimensional plots, trends in the samples during the process, or detect abnormal data (outliers).

In this case, the data matrix is organized with samples in the rows and wavelengths in the columns, and in case of several batch samples the time-wise unfolding (*IK*x*J*) is employed. Prior to unfolded PCA, the data matrix is column mean centred, so the average considers every sample and time.

The main changes during alcoholic fermentation are usually reflected in the score values of the first PC, as it contains the maximum variability of the samples. It has been shown

by Buratti *et al.*[17] that the first PC plotted versus time fits the fermentation kinetic evolution (Equation 1), as the general form of the Gompertz equation.

$$C = C_{\infty} \cdot exp\left\{-exp\left[\frac{\mu_{max} \cdot e}{C_{\infty}}(\lambda - t) + 1\right]\right\} \qquad \text{Equation 1}$$

where μ_{max} is the representation of maximum specific growth rate, λ the length of lag phase, C_{∞} the curve asymptote, depending on the compounds considered, ethanol or sugars and C is the concentration at time t or the concentration of sugars consumed.

A similar trend is observed by visual inspection of the sigmoid functions of the first PC versus time and density versus time [see e.g. Burati *et al.*[18], Mehmood *et al.*[34] and Figure 1]. This behavior is confirmed by a good mathematical correlation between the scores of the first PC and the density values.

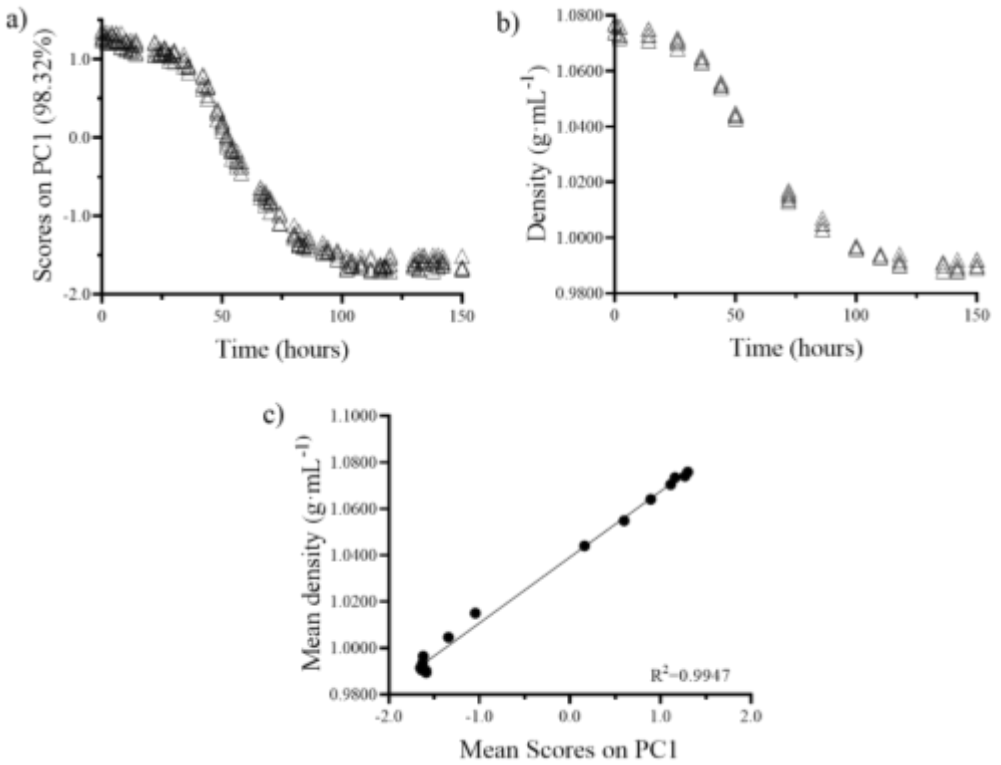


Figure 1. PC1 score values plotted versus time (a), density plotted versus time (b) during alcoholic fermentation and mean PC1 score values plotted versus mean density (c).

To obtain the evolution of the scores over time shown in figure 1a, we followed 5 alcoholic fermentations inoculated with $3 \cdot 10^6$ CFU $\cdot$ mL^{-1} of *Saccharomyces cerevisiae*. The dimension of the data matrix was (5x845x48), which was time-wise unfolded into a matrix of (240x845). The unfolded matrix was pre-processed using a 15–points smoothing, then SNV and finally column mean centering. After this, the PCA model was built and the trend observed in the PC1 score values clearly recalls the evolution of density during the fermentation process, as can be seen on the graph of density measurements (Figure 1b) throughout the process. Figure 1c shows that a good linear correlation (> 0.99%) is obtained between the mean values of the scores on first PC and the mean density for the times in which both measurements were carried out. After PCA modelling new samples can be qualitative evaluated with this methodology. For that purpose, the same data unfolding is performed, data are centred using the mean of the calibration data and new samples are finally projected onto the PCA model.

While PCA returns qualitative information, Multivariate Statistical Process Control (MSPC) evaluates, with statistical limits, whether the process is running under control. The scores and the residuals of the PCA model are used to build Hotelling T^2 and Q charts, respectively[35].

To reproduce the process conditions in a small-scale laboratory experiment, different fermentation batches must be monitored, and samples measured over time following the rules mentioned above. It is also possible to consider additional sources of variability, such as the initial sugar concentration or the ripening state of the grapes. A PCA model is then built, in which scores (from one or more PCs) and residuals are used as parameters to define the statistical limits (samples are considered to belong to a normal distribution) of the control charts (T^2 and Q) and to monitor them for future batches.

Once the model has been built, and the statistical limits have been calculated, new samples from an ongoing process are projected onto the model to determine whether the process is running under normal operation conditions (NOC). For that, the projected samples must have the same spectral range and must have been pre-treated in the same way as the modelling samples. New data are centred with the mean obtained from calibration data. Literature shows a wide range of applications and types of control charts[18,36,37] and references therein; for example, Cavaglia *et al*. have shown that building control charts with various statistical parameters (PC1 and PC2 scores values[18], Q residuals and Hoteling T^2 [34], or evolution over time) allows detecting unwanted

subprocesses, such as malolactic fermentation. Malolactic fermentation is related to the presence of unwanted microorganisms, lactic acid bacteria, which exist in the ecology of the vineyard and also in the winery[38]. These microorganisms, like other contaminants (yeasts and acetic acid bacteria), produce their metabolites in low amounts, making it difficult to detect them using the whole spectral range. This is the reason why, when it comes to suspected contamination (mainly when the tracked parameters are in the minority), it is necessary to focus on the specific region of the substrate, the product or both[18,39]. An example of this application is shown in Figure 2.

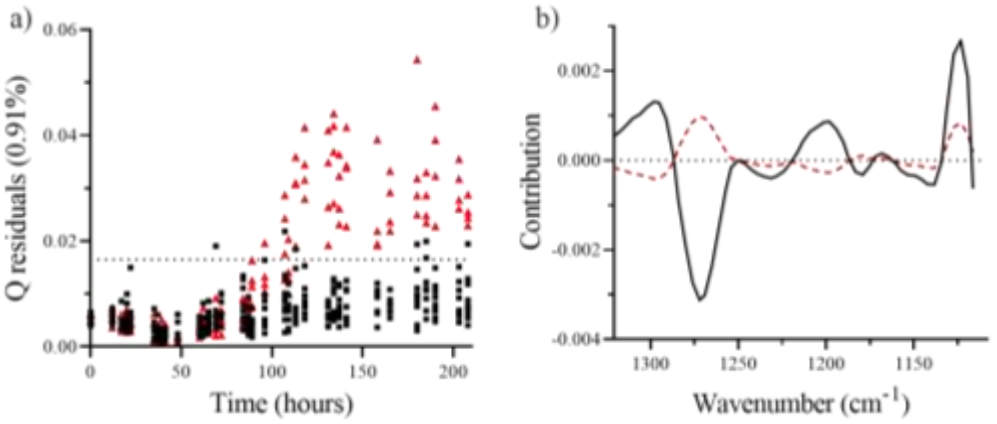


Figure 2. PCA based control chart (left) and the mean contribution plot (right) for normal process samples (squares) and deviated process samples (triangles and dashed line).

Figure 2 shows the evolution over time of Q residuals (from a PCA model built with 2 PCs) obtained from 10 normal alcoholic fermentations and 4 fermentations intentionally contaminated with lactic acid bacteria at a concentration of $2.5 \cdot 10^6$ CFU $\cdot$ mL^{-1}. The best PCA model was obtained when using specifically the region of organic acids from 1250 to 1089 cm^{-1}. It may be explained as the metabolism of lactic acid bacteria consists of converting malic acid into lactic acid which absorb in that specific region. The data matrix obtained was of dimension (14x62x34) and it was time-wise unfolded into an array of (506x62). After unfolding, the data matrix was preprocessed by 15–points smoothing, SNV and all fermentations were mean centered.

To detect process deviations, PCA-based control charts can be applied. However, to meet the process control requirements and assigning the cause of the deviation the contribution plots are required[35]. Contribution plots show the main regions of the spectra that are used

to differentiate normal and abnormal process samples, as shown in Figure 2b. Contribution plots are obtained as the multiplication of the loadings by the spectra, which increases the information of the spectra related to specific variables. Therefore, for a certain spectral region, deviations related to specific compounds can be detected. As it can be seen in the contribution plot (right), each type of fermentation showed a different trend.

5.3. Prediction of physicochemical parameters

As the IR spectrum contains information about the main species present in the sample, it is possible to correlate the changes in the IR spectrum with the evolution of the process, i.e. changes in the reacting species. For this purpose, Partial Least Square (PLS) regression is usually applied. The PLS regression method works like PCA, compressing the spectral information into a few latent variables, but also correlating this spectral information to a **Y** matrix that contains the values of a physicochemical parameter to be predicted. Typical oenological parameters in wineries are density, pH, volatile acidity, total acidity, and initial YAN concentration, but it has been shown that is also possible to predict other properties, such as organoleptic attributes, polyphenols and other minor compounds[25].

To build a PLS model, it is necessary to use a reference or standard analytical method to estimate the y values, in the same way that the use of standards is required in many other analytical methods. Therefore, to monitor and control the conventional parameters, it is necessary to analyze at the early stages every sample with both methods, the standard and the infrared.

To build the PLS model a time-wise unfolding (IKxJ) is necessary to statistically correlate the spectrum at a given time with the measured property. This also means that the samples must have been analyzed by the reference methods of analysis, which often are time consuming. This limiting factor implies a decrease in the number of points that can be taken to obtain correlation between a property and the spectrum, as it is illustrated in Figure 1.

When a new sample is analyzed, its similarity to the calibration data is evaluated through the model, which predicts the parameter value. Projected data are spectrally preprocessed as the calibration samples and centered using the mean of the calibration data. The prediction ability of the calibration models is assessed by the Root Mean Square Error of

Prediction (RMSEP) or standard error of prediction (SEP). In addition, when the number of samples is not enough to split the samples into calibration and validation sets, it is possible to evaluate the prediction of the model through Root Mean Square Error of Cross-Validation (RMSECV). It is important to note that an averaged error (along the whole range) is provided by this methodology[25].

The prediction of wine parameters has been extensively studied by varying vintage and grape varieties in order to cover typical oenological parameter ranges[40]. This methodology should be considered in the context of process control, to obtain robust models, since the samples must cover the full range of variation in a given process and also between processes with different initial characteristics[41]. The prediction of sugars (glucose and fructose) allows building a control chart to plot this parameter versus time. This has been shown by several authors, obtaining good prediction errors: 5 $g \cdot L^{-1}$ and 2 $g \cdot L^{-1}$ at the end of the fermentation[41], or even 10.9 $g \cdot L^{-1}$ with a portable ATR–MIR[22]. Other parameters typically predicted when monitoring the alcoholic fermentation are density and pH. They have been predicted using a portable ATR–MIR with an error of prediction of 0.0014 $g \cdot mL^{-1}$ and 0.07, respectively[18]. Other important parameters, such as phenolic compounds, anthocyanins and flavonoids, which are important in red wine alcoholic process control, have also been predicted using unfolded PLS regression[42].

It is also possible to build a control chart through the prediction of a property using the unfolded PLS algorithm, and to establish a critical limit, from a legal or quality perspective, to decide what samples are under or out-of-control (Figure 3).

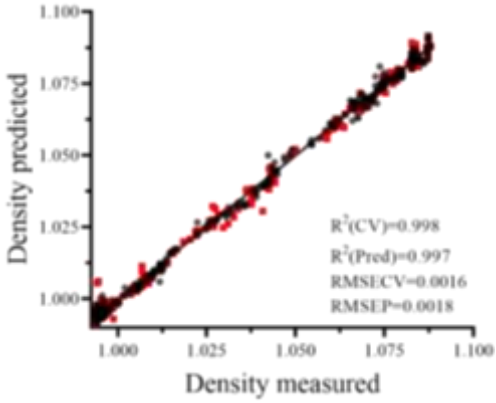


Figure 3. PLS model to predict density ($g \cdot mL^{-1}$) along alcoholic fermentation.

The above mentioned unfolded PLS regression model (shown in Figure 3) is an example of the density prediction ability of ATR–MIR. The example uses a **X** calibration data matrix of (580x845) that was correlated to a **Y** matrix of (580x1). As it can be seen both matrixes have the same rows as each sample used for prediction must be analyzed by standard physicochemical analysis, in this case by a portable densimeter. Also, in Figure 3, **X** test data matrix of (528x845) that also has an **Y** matrix of (528x1) were used to stablish prediction error. The prediction ability of the models is deeply dependent on the number of samples used, as more samples means a better estimation of the sample variability. Moreover, having more samples allow to improve the validation of the models.

It is also possible to detect process deviations if some compounds, such as acetic acid, are predicted. Acetic acid is a compound generated by yeast in order to obtain energy in sub-optimal process conditions, such as YAN shortfall or when there is an increase in process temperature. Urtubia *et al*. predicted sugars, ethanol, glycerol, succinic and acetic acids and showed that the small errors obtained made possible to detect the miss-behaviour from a temperature-gradient fermentation and a YAN shortfall fermentation[43].

Prediction of fermentation evolution

The spectral information can also be used to predict the time points of the fermentation as alcoholic fermentation is a bioprocess evolving over time. Cozzolino *et al*. have showed that it is possible to predict the time course of a wild fermentation, with an error of 1.21 days[44]. This methodology would be applicable to detect stuck and sluggish fermentations, as the predicted value of time course of fermentation would be smaller than the actual one. From a practical oenological point of view, an error of more than one day is not a suitable solution. Therefore, another methodology, such as the biological time, is needed to decrease the error.

Biological time, firstly introduced by Jorgensen *et at*.[45], allows detecting slight differences in the fermentation process due to the metabolism of yeasts. First, the data is relativized in a scale from 0 to 1, afterwards PLS regression is applied, and the calibration equation is used to predict the time of the spectra used. This circular approach allows determining the biological time, due to small differences in sugar consumption along fermentations. Then, PLS is re-applied with the predicted biological time. The scores convey into an alcoholic fermentation control chart and a 95% confidence limits are calculated. The

calibration model will be reused in future fermentations and the scores will be plotted in the control chart to determine how normal the time involved is.

Cavaglia *et al*. have shown that the biological time approach described above provides good results to detect fermentations with small yeast assimilable nitrogen (YAN) concentration. The control chart showed that a YAN fermentation was in the 0.6 of the biological time at the time it should be 1[22].

As shown in Figure S2, through biological time it is possible to obtain the confidence intervals of PLS factor 1. The dimension of the spectral data matrix was (20x845x9) and it was time-wise unfolded into a matrix of (180x845). The unfolded spectral matrix was preprocessed using an 11–points smoothing, then SNV and finally column mean centered. The **Y1** matrix (180x1) that contained the times of the alcoholic fermentation process was relativized from 0 to 1. The predicted parameters were obtained in the **Y2** matrix (180x1). The prediction was re-built with the **Y2** matrix, obtaining the biological time in the **Y3** matrix. Besides, the confidence intervals were calculated with the values of PLS factor 1 using **Y3**[22,45]. A new spectral data matrix (27x845) of 3 nitrogen deficient alcoholic fermentations was projected onto the model to predict their biological time, and its representation (Figure S2) showed slower fermentation kinetics.

Detection of deviations from NOC

To determine whether the samples are under or out-of-control, it is possible to apply PLS–DA. In PLS–DA models, a regression is performed between the spectra (**X** matrix) and a **y**-vector containing a dichotomous variable (typically 0 and 1) that expresses the type of process, under and out-of-control, respectively. This methodology can improve the detection of deviations, as low concentrated compounds, which are under the quantification threshold, could not be predicted but the overall spectral changes are sufficient to detect deviation. With this methodology, local and batch-wise unfolding approaches may be used.

Local unfolding is used to determine if the behaviour of a sample in a specific time or time gap is under-control. If the sampling pattern is reproducible over the different measured processes, that is, if sampling is always performed at the same time during the fermentation with a given time interval (e.g. each *k* hours), a local *k* time unfolding is used summoning the same time for all the samples. In wineries, the sampling pattern may not

be standard over years, so in order to use historical data it is necessary to use a gap time for the models. We have shown that an 8–hour approach improves process control and allows corrective actions to be taken[34].

To increase the classification performance, it is even possible to apply a moving window approach, firstly introduce by Camacho *et al.*[46]. In this approach, a *k* number of times are gathered, and PLS–DA is applied (from time *n* to time *n+k*). The models are built by moving the times used, from time *n+1* to time *n+k+1*. This approach is partially batch-wise (*IxJK*) unfolded, where *K* is a subset of *K*. Besides, using the batch-wise unfolding the result will express if the whole process is under NOC conditions.

The performance evaluation of the PLS–DA models is done using a discrimination threshold (between 0 and 1), which is calculated from the probability of classification error of the samples in the classes. The discrimination threshold is calculated taking into account a Gaussian distribution of the predicted classes, and the y value at which the two curves converge is the discrimination threshold[47] (see blue and red dashed lines in Figure 4). The optimization of the discrimination threshold is based on the Receiver Operating Characteristic (ROC) curve, which is a graphical representation of the specificity and sensitivity variation as a function of the threshold. Specificity is the ratio of true positives to the total of true positives plus false negatives; and selectivity is the ratio of true negatives to the total of true negatives plus false positives Cavaglia *et al.* have shown that local approaches to distinguish YAN shortfall[22] and lactic acid bacteria contamination[18] provided 100% correct classification of the classes. However, when applying the moving window approach in a YAN shortfall process, the lack of sampling points did not show better results[22].

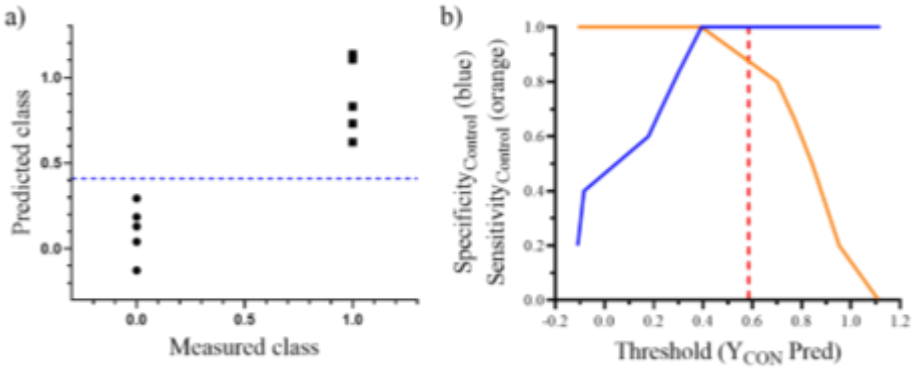


Figure 4. a) PLS-DA model for under control (zero) and out-of-control (one) samples. b) Receiver Operating Characteristic curve for the under-control class.

The PLS–DA model shown in Figure 4a was built using 5 normal alcoholic fermentations and 5 fermentations that were intentionally contaminated by co-inoculating a population of $1 \cdot 10^6$ CFU $\cdot$ mL^{-1} lactic acid bacteria, to reproduce typical contamination at the start of the alcoholic fermentation. Data used in this k model is from 213 hours, as in this point the 50% of malic acid was consumed. Data matrix was (10x62), as the model focused only in the organic acids' region from 1250 to 1089 cm^{-1}.

Figure 4b shows the ROC curve for the under control class. The threshold for under control samples is 1-threshold for out-of-control samples. This is why the threshold is calculated to maximize both sensitivity and selectivity for the two categories. In the example, under control class threshold is 0.5849 and for out-of-control class is 0.4151 (which is the one showed in Figure 4a). In the example, specificity for both classes in the optimized thresholds is 1; and selectivity is 0.925 for out-of-control class (data not shown) and 0.875 for control class.

Conclusions

This tutorial provides practical methodologies that can be used to study wine alcoholic fermentation at a laboratory microscale with ATR–FT–MIR spectroscopy and following PAT recommendations. The speed and portability of current FT–MIR equipment are characteristics that wineries demand, so this technique can have a wide application in this field. In that sense, many of the given recommendations can be extended to other vibrational techniques.

It should be noted that the combination of spectroscopy with chemometrics allows to obtain many features, almost at the same time, from the same dataset such as the monitoring of the fermentation, the prediction of relevant parameters and even the detection of deviations.

Finally, this tutorial can help the oenological researcher, who is not usually familiar with chemometrics, not to get lost among all the chemometric approaches available in the literature. Thus, this report explains the chemometric techniques, without deepening into chemometric algorithms but keeping the scientific rigor, which can be applied in each case followed by a discussion of the results that can be obtained.

Funding

This project has been possible with the support of the Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación (AEI) (project PID2019-104269RR-C33); and with the support of the Ministerio de Ciencia, Innovación y Universidades, the AEI and the European Social Fund (FEDER) (project AGL2015-70106-R). This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn−Garcia, D.). Cavaglia, J. would also like to thank the Catalan Agency for Management of University and Research Grants (AGAUR) for the FI grant (FI_B100154).

References (Paper 1)

1. Ribéreau-Gayon, P.; Dubourdieu, D.; Donèche, B.; Lonvaud, A. Handbook of Enology Volume 1. The Microbiology of Wine and Vinifications; 2nd Editio.; **2006**; ISBN 0-470-01034-7.

2. Urtubia, A.; Hernández, G.; Roger, J.M. 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.

3. 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, 282–297, doi:10.1080/05704928.2015.1132721.

4. Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacuring, and Quality Assurance; Services, U.S. Department of Health and Human, Food and Drug Administration, **2004**.

5. Jerome, R.E.; Singh, S.K.; Dwivedi, M. Process Analytical Technology for Bakery Industry: A Review. *J Food Process Eng* **2019**, 42, 1–21, doi:10.1111/jfpe.13143.

6. van den Berg, F.; Lyndgaard, C.B.; Sørensen, K.M.; Engelsen, S.B. Process Analytical Technology in the Food Industry. *Trends Food Sci Technol* **2013**, 31, 27–35, doi:10.1016/j.tifs.2012.04.007.

7. Pu, Y.Y.; O'Donnell, C.; Tobin, J.T.; O'Shea, N. Review of Near-Infrared Spectroscopy as a Process Analytical Technology for Real-Time Product Monitoring in Dairy Processing. *Int Dairy J* **2020**, 103, 104623, doi:10.1016/j.idairyj.2019.104623.

8. Lourenço, N.D.; Lopes, J.A.; Almeida, C.F.; Sarraguça, M.C.; Pinheiro, H.M. Bioreactor Monitoring with Spectroscopy and Chemometrics: A Review. *Anal Bioanal Chem* **2012**, 404, 1211–1237, doi:10.1007/s00216-012-6073-9.

9. Porep, J.U.; Kammerer, D.R.; Carle, R. On-Line Application of near Infrared (NIR) Spectroscopy in Food Production. *Trends Food Sci Technol* **2015**, 46, 211–230, doi:10.1016/j.tifs.2015.10.002.

10. Grassi, S.; Alamprese, C. Advances in NIR Spectroscopy Applied to Process Analytical Technology in Food Industries. *Curr Opin Food Sci 2018*, 22, 17–21, doi:10.1016/j.cofs.2017.12.008.

11. Sun, D.W. Infrared Spectroscopy for Food Quality Analysis and Control; Sun, D.W., Ed.; 1st Ed.; Elsevier, **2009**; ISBN 978-0-12-374136-3.

12. International Organization of vine and wine (OIV) International Oenological Codex; **2019**; ISBN 9782850380013.

13. Aranda, A.; Matallana, E.; del Olmo, M. Saccharomyces Yeasts I: Primary Fermentation. In *Molecular Wine Microbiology*; **2011**; pp. 1–31 ISBN 9780123750211.

14. International Organization of vine and wine (OIV) Chemical Acidification of Musts. In *International Code of Oenological Practices*; **2015**; p. II.2.1-8.

15. Bell, S.J.; Henschke, P.A. Implications of Nitrogen Nutrition for Grapes, Fermentation and Wine. *Aust J Grape Wine Res* **2005**, 11, 242–295, doi:10.1111/j.1755-0238.2005.tb00028.x.

16. Padilla, B.; Gil, J. V.; Manzanares, P. Past and Future of Non-*Saccharomyces* Yeasts: From Spoilage Microorganisms to Biotechnological Tools for Improving Wine Aroma Complexity. *Front Microbiol* **2016**, 7, 1–20, doi:10.3389/fmicb.2016.00411.

17. Buratti, S.; Ballabio, D.; Giovanelli, G.; Dominguez, C.M.Z.; Moles, A.; Benedetti, S.; Sinelli, N. 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.

19. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

20. Subramanian, A.; Rodriguez-Saona, L. Fourier Transform Infrared (FTIR) Spectroscopy; 1st ed.; Elsevier Inc., 2009; Vol. 1; ISBN 9780123741363.

21. Crocombe, R.A. Portable Spectroscopy. *Appl Spectrosc* **2018**, 72, 1701–1751, doi:10.1177/0003702818809719.

22. Gorla, G.; Mestres, M.; Boqué, R.; Riu, J.; Spanu, D.; Giussani, B. ATR-MIR Spectroscopy to Predict Commercial Milk Major Components: A Comparison between a Handheld and a Benchtop Instrument. *Chemometr Intell Lab* **2020**, 200, 103995, doi:10.1016/j.chemolab.2020.103995.

23. Cavaglia, J.; Giussani, B.; Mestres, M.; Puxeu, M.; Busto, O.; Ferré, J.; Boqué, R. Early Detection of Undesirable Deviations in Must Fermentation Using a Portable FTIR-ATR Instrument and Multivariate Analysis. *J Chemom* **2019**, 33, e3162, doi:10.1002/cem.3162.

24. Cozzolino, D. Sample Presentation, Sources of Error and Future Perspectives on the Application of Vibrational Spectroscopy in the Wine Industry. *J Sci Food Agric* **2015**, 95, 861–868, doi:10.1002/jsfa.6733.

25. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

26. dos Santos, C.A.T.; Páscoa, R.N.M.J.; Lopes, J.A. A Review on the Application of Vibrational Spectroscopy in the Wine Industry: From Soil to Bottle. *TRAC Trends Anal Chem* **2017**, 88, 100–118, doi:10.1016/j.trac.2016.12.012.

27. Rinnan, Å.; Norgaard, L.; van den Berg, F.; Thygesen, J.; Bro, R.; Engelsen, S.B. Data Pre-Processing. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Elsevier Ltd, 2009; pp. 29–50 ISBN 978-0-12-374136-3.

28. Wu, Z.; Xu, E.; Long, J.; Zhang, Y.; Wang, F.; Xu, X.; Jin, Z.; Jiao, A. Monitoring of Fermentation Process Parameters of Chinese Rice Wine Using Attenuated Total Reflectance Mid-Infrared Spectroscopy. *Food Control* **2015**, 50, 405–412, doi:10.1016/j.foodcont.2014.09.028.

29. Rinnan, Å.; Berg, F. van den; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TRAC - Trends Anal Chem* **2009**, 28, 1201–1222, doi:10.1016/j.trac.2009.07.007.

30. Kumar, N.; Bansal, A.; Sarma, G.S.; Rawal, R.K. Chemometrics Tools Used in Analytical Chemistry: An Overview. *Talanta* **2014**, 123, 186–199, doi:10.1016/j.talanta.2014.02.003.

31. Berrueta, L.A.; Alonso-Salces, R.M.; Hberger, K. Supervised Pattern Recognition in Food Analysis. *J Chromatogr A* **2007**, 1158, 196–214, doi:10.1016/j.chroma.2007.05.024.

32. Ballabio, D.; Grisoni, F.; Todeschini, R. Multivariate Comparison of Classification Performance Measures. *Chemometr Intell Lab* **2018**, 174, 33–44, doi:10.1016/j.chemolab.2017.12.004.

33. Westad, F.; Marini, F. Validation of Chemometric Models - A Tutorial. *Anal Chim Acta* **2015**, 893, 14–24, doi:10.1016/j.aca.2015.06.056.

34. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A Review of Variable Selection Methods in Partial Least Squares Regression. *Chemometr Intell Lab* **2012**, 118, 62–69, doi:10.1016/j.chemolab.2012.07.010.

35. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Monitoring Wine Fermentation Deviations Using an ATR-MIR Spectrometer and MSPC Charts. *Chemometr Intell Lab* **2020**, 201, 104011, doi:10.1016/j.chemolab.2020.104011.

36. Bersimis, S.; Panaretos, J.; Psarakis, S. Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry. *Qual Reliab Eng Int* **2009**, 23, 517–543, doi:10.48550/arXiv.0901.2880.

37. Catelani, T.A.; Santos, J.R.; Páscoa, R.N.M.J.; Pezza, L.; Pezza, H.R.; Lopes, J.A. Real-Time Monitoring of a Coffee Roasting Process with near Infrared Spectroscopy Using Multivariate Statistical Analysis: A Feasibility Study. *Talanta* **2018**, 179, 292–299, doi:10.1016/j.talanta.2017.11.010.

38. Ávila, T.C.; Poppi, R.J.; Lunardi, I.; Tizei, P.A.G.; Pereira, G.A.G. Raman Spectroscopy and Chemometrics for On-Line Control of Glucose Fermentation by *Saccharomyces cerevisiae*. *Biotechnol Prog* **2012**, 28, 1598–1604, doi:10.1002/btpr.1615.

39. Vaudano, E.; Costantini, A.; Garcia-Moruno, E. Microbial Ecology of Wine. In *Modeling the Microbial Ecology of Foods: Quantitative Microbiology in Food Processing*; **2017**; pp. 547–559 ISBN 9781118823071.

40. Grassi, S. Microbial Food Fermentations: Innovative Approach Using Infrared Spectroscopy, Università degli Studi di Milano. Milan (Italy), **2013**.

41. 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.

42. Bellon, V. Fermentation Control Using ATR and an FT-IR Spectrometer. *Sens Actuators B Chem* **1993**, 12, 57–64, doi:10.1016/0925-4005(93)85015-3.

43. 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, 947–955, doi:10.1007/s00217-010-1227-5.

44. Urtubia, A.; Pérez-correa, J.R.; 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.

45. 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.

46. Jørgensen, P.; Pedersen, J.G.; Jensen, E.P.; Esbensen, K.H. On-Line Batch Fermentation Process Monitoring (NIR) - Introducing "Biological Process Time". *J Chemom* **2004**, 18, 81–91, doi:10.1002/cem.850.

47. Camacho, J.; Picó, J.; Ferrer, A. The Best Approaches in the On-Line Monitoring of Batch Processes Based on PCA: Does the Modelling Structure Matter? *Anal Chim Acta* **2009**, 642, 59–68, doi:10.1016/j.aca.2009.02.001.

48. Pérez, N.F.; Ferré, J.; Boqué, R. Calculation of the Reliability of Classification in Discriminant Partial Least-Squares Binary Classification. *Chemometr Intell Lab* **2009**, 95, 122–128, doi:10.1016/j.chemolab.2008.09.005.

Supplementary Material

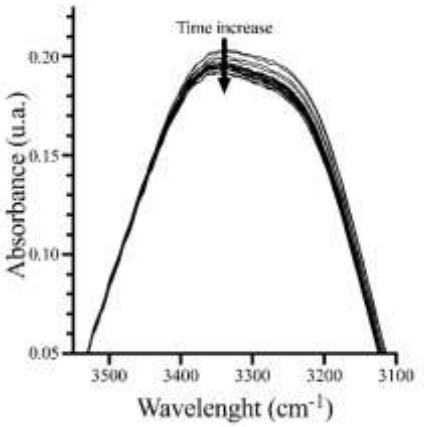


Figure S1. Evolution over time of the region of the water band (3000 – 3750 cm^{-1}) of the same must sample from the switching-on of the spectrometer until the equilibration time.

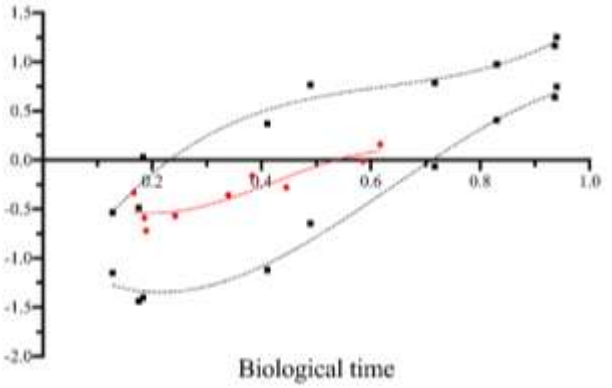


Figure S2. Control charts based on biological time. 95% confidence intervals (dashed lines) have been built from the scores of the NOC samples (average ± 2s for PLS factor 1). Adapted from Cavaglia *et al.*[23], with author's permission policy of Wiley.

References (Introduction)

1. Hystory of Wine. In *The Chemistry and Biology of Winemaking*; Hornsey, I., Ed.; The Royal Society of Chemistry: Cambridge, **2007**; pp. 11–64.

2. Estreicher, S.K. A Brief History of Wine in Spain. *European Review* **2000**, 21, 209–239, doi:10.1017/s1062798712000373.

4. Carrasco, I.; Castillo-Valero, J.S.; Córcoles, C.; Carchano, M. Greening Wine Exports? Changes in the Carbon Footprint of Spanish Wine Exports. *Int J Environ Res Public Health* **2021**, 18(17), 9035, doi:10.3390/ijerph18179035.

5. International Organization of Grape and Wine. The World Organic Vineyard. In *Focus OIV*; Dijon, **2021**; pp. 1–21.

6. International Organization of Grape and Wine. State of the World Vine and Wine Sector 2021; Dijon, **2022**; pp. 1–20.

7. International Organization of Grape and Wine. World Wine Production Outlook. In *OIV First Estimates*; Dijon, **2022**; pp. 1–8.

8. Reynier, A. Ampelografía y Mejora de Variedades. In *Manual de Viticultura*; Reynier, A., Ed.; Ediciones Omega: Barcelona, **2013**; pp. 3–84.

9. Blouin, J.; Guimberteau, G. Maduración y Madurez de La Uva; Blouin, J., Guimberteau, G., Eds.; Ediciones Mundi-Prensa: Madrid, **2004**.

10. Reynier, A. Fisiología de La Vid. In *Manual de viticultura*; Ediciones Omega: Barcelona, **2013**; pp. 117–155.

11. Hidalgo Togores, J. Morfología, Maduración y Composición Del Racimo. In *Tratado de Enología*; Ediciones Mundi-Prensa: Madrid, **2010**; pp. 57–222.

12. Kliewer, W.M. The Glucose-Fructose Ratio of *Vitis vinifera* Grapes. *Am J Enol Vitic* **1967**, 18, 33–41, doi:10.5344/ajev.1967.18.1.33.

13. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

14. Sweetman, C.; Deluc, L.G.; Cramer, G.R.; Ford, C.M.; Soole, K.L. Regulation of Malate Metabolism in Grape Berry and Other Developing Fruits. *Phytochemistry* **2009**, 70, 1329–1344, doi:10.1016/j.phytochem.2009.08.006.

15. Kennedy, J.A. Grape and Wine Phenolics: Observations and Recent Findings. *Cienc Investig Agrar* **2008**, 35(2), 107–120, doi:10.4067/s0718-16202008000200001.

16. Monagas, M.; Bartolomé, B.; Gómez-Cordovés, C. Updated Knowledge About the Presence of Phenolic Compounds in Wine. *Crit Rev Food Sci Nutr* **2007**, 45, 85–118, doi:10.1080/10408690490911710.

17. Barceló, A.R.; Calderón, A.A.; Zapata, J.M.; Muñoz, R. The Histochemical Localization of Anthocyanins in Seeded and Seedless Grapes (*Vitis vinifera*). *Sci Hortic* **1994**, 57, 265–268, doi:10.1016/0304-4238(94)90146-5.

18. Souquet, J.M.; Labarbe, B.; Le Guernevé, C.; Cheynier, V.; Moutounet, M. Phenolic Composition of Grape Stems. *J Agric Food Chem* **2000**, 48, 1076–1080, doi:10.1021/jf991171u.

19. Prieur, C.; Rigaud, J.; Cheynier, V.; Moutounet, M. Oligomeric and Polymeric Procyanidins from Grape Seeds. *Phytochemistry* **1994**, 36, 781–784, doi:10.1016/s0031-9422(00)89817-9.

20. Martínez de Toda, F. Efecto de Las Altas Temperaturas Sobre La Vid, La Uva y El Vino. In *Técnicas vitícolas frente al cambio climático*; Hernández Úbeda, I., Ed.; Ediciones Mundi-Prensa: Santiago de Compostela, **2019**. ISBN 9788484767565.

21. Gutiérrez-Gamboa, G.; Zheng, W.; Martínez de Toda, F. Current Viticultural Techniques to Mitigate the Effects of Global Warming on Grape and Wine Quality: A Comprehensive Review. *Food Res Int* **2021**, 139, 109946, doi:10.1016/j.foodres.2020.109946.

22. Pons, A.; Allamy, L.; Schüttler, A.; Rauhut, D.; Thibon, C.; Darriet, P. What Is the Expected Impact of Climate Change on Wine Aroma Compounds and Their Precursors in Grape? *OENO One* **2017**, 51, 141–146, doi:10.20870/oeno-one.2017.51.2.1868.

23. Hjelmeland, A.K.; Ebeler, S.E. Glycosidically Bound Volatile Aroma Compounds in Grapes and Wine: A Review. *Am J Enol Vitic* **2015**, 66, 1–11, doi:10.5344/ajev.2014.14104.

24. Ugliano, M.; Henschke, P.A. Yeasts and Wine Flavour. In *Wine Chemistry and Biochemistry*; Moreno-Arribas, M.V., Polo, M.C., Eds.; Springer: Madrid, **2009**; pp. 313–392. doi: 10.1007/978-0-387-74118-5_17.

25. Reynier, A. Razonamiento y Organización de La Recolección. In *Manual de Viticultura*; Reynier, A., Ed.; Ediciones Omega: Barcelona, **2013**; pp. 389–406.

26. Iland, P.; Bruer, N.; Ewart, A.; Markides, A.; Sitters, J. Monitoring the Winemaking Process from Grapes to Wine: Techniques and Concepts; 2nd Edition.; Patrick Iland Wine Promotions Pty LTD: Adelaide, **2012**. ISBN: 978-0-9581605-6-8.

27. Evaluation by Refractometry of the Sugar Concentration in Grape Musts, Concentrated Grape Musts and Rectified Concentrated Grape Musts - Method OIV-MA-AS2-02. In *Compendium of international methods of analysis of wines and musts*; International Organization of vine and wine (OIV): Dijon, **2022**.

28. pH – Method OIV-MA-AS313-15. In *Compendium of international methods of analysis of wines and musts*; International Organization of vine and wine (OIV): Dijon, **2022**.

29. Casalta, E.; Sablayrolles, J.M.; Salmon, J.M. Comparison of Different Methods for the Determination of Assimilable Nitrogen in Grape Musts. *LWT - Food Sci Technol* **2013**, 54, 271–277, doi:10.1016/j.lwt.2013.05.009.

30. Bourzeix, M.; Dubernet, M.; Heredia, N. Sur l'extraction de Divers Constituants Phénoliques Des Raisins et de Leur Rafles. *Industries Alimentaires et Agricoles* **1975**, 9, 1057–1064.

31. Iland, P.G.; Cynkar, W.; Francis, I.L.; Williams, P.J.; Coombe, B.C. Optimisation of Methods for the Determination of Total and Red-Free Glycosyl Glucose in Black Grape Berries of *Vitis vinifera*. *Aust J Grape Wine Res* **1996**, 2, 171–178, doi:10.1111/j.1755-0238.1996.tb00105.x.

32. Gunata, Y.Z.; Bayonove, C.L.; Baumes, R.L.; Cordonnier, R.E. The Aroma of Grapes I. Extraction and Determination of Free and Glycosidically Bound Fractions of Some Grape Aroma Components. *J Chromatogr A* **1985**, 331, 83–90, doi:10.1016/0021-9673(85)80009-1.

33. Hornsey, I. Winemaking Processes. In *The Chemistry and Biology of Winemaking*; The Royal Society of Chemistry: Cambridge, **2007**; pp. 161–202.

34. Waterhouse, A.L.; Sacks, G.L.; Jeffery, D.W. Understanding Wine Chemistry; Wiley, **2016**. doi:10.1002/9781118730720.

35. Kunkee, R.E. Some Roles of Malic Acid in the Malolactic Fermentation in Wine Making. *FEMS Microbiol Lett* **1991**, 88, 55–71, doi:10.1016/0378-1097(91)90696-8.

36. Muñoz, R.; Moreno-Arribas, M.V.; de las Rivas, B. Lactic Acid Bacteria. In *Molecular Wine Microbiology*. Carrascosa, A.V.; Muñoz, R.; González, R. Ed. **2011**, 191–226, doi:10.1016/b978-0-12-375021-1.10008-6.

37. Martínez-Gil, A.; del Alamo-Sanza, M.; Sánchez-Gómez, R.; Nevares, I. Different Woods in Cooperage for Oenology: A Review. *Beverages* **2018**, 4(4), 94, doi:10.3390/beverages4040094.

38. Carpena, M.; Pereira, A.G.; Prieto, M.A.; Simal-Gandara, J. Wine Aging Technology: Fundamental Role of Wood Barrels. *Foods* **2020**, 9, 1160, doi:10.3390/foods9091160.

39. Albergaria, H.; Arneborg, N. Dominance of *Saccharomyces cerevisiae* in Alcoholic Fermentation Processes: Role of Physiological Fitness and Microbial Interactions. *Appl Microbiol Biotechnol* **2016**, 100, 2035–2046, doi:10.1007/s00253-015-7255-0.

40. Portillo, M. del C.; Franquès, J.; Araque, I.; Reguant, C.; Bordons, A. Bacterial Diversity of Grenache and Carignan Grape Surface from Different Vineyards at Priorat Wine Region (Catalonia, Spain). *Int J Food Microbiol* **2016**, 219, 56–63, doi:10.1016/j.ijfoodmicro.2015.12.002.

41. Wibowo, D.; Eschenbruch, R.; Davis, C.R.; Fleet, G.H.; Lee, T.H. Occurrence and Growth of Lactic Acid Bacteria in Wine: A Review. *Am J Enol Vitic* **1985**, 36, 302–313, doi:10.5344/ajev.1985.36.4.302.

42. Barnett, J.A.; Entian, K.D. A History of Research on Yeasts 9: Regulation of Sugar Metabolism. *Yeast* **2005**, 22, 835–894, doi:10.1002/yea.1249.

43. Ingledew, W.M.; Magnus, C.A.; Sosulski, F.W. Influence of Oxygen on Proline Utilization During the Wine Fermentation. *Am J Enol Vitic* **1987**, 38, 246–248, doi:10.5344/ajev.1987.38.3.246.

44. Martínez-Moreno, R.; Morales, P.; Gonzalez, R.; Mas, A.; Beltran, G. Biomass Production and Alcoholic Fermentation Performance of *Saccharomyces cerevisiae* as a Function of Nitrogen Source. *FEMS Yeast Res* **2012**, 12, 477–485, doi:10.1111/j.1567-1364.2012.00802.x.

45. Hornsey, I. The Yeast and Fermentation. In *The Chemistry and Biology of Winemaking*; The Royal Society of Chemistry: Cambridge, **2007**; pp. 114–160. ISBN: 978-1-84755-766-7.

46. Gobert, A.; Tourdot-Maréchal, R.; Sparrow, C.; Morge, C.; Alexandre, H. Influence of Nitrogen Status in Wine Alcoholic Fermentation. *Food Microbiol* **2019**, 83, 71–85, doi:10.1016/j.fm.2019.04.008.

47. Albergaria, H.; Arneborg, N. Dominance of *Saccharomyces cerevisiae* in Alcoholic Fermentation Processes: Role of Physiological Fitness and Microbial Interactions. *Appl Microbiol Biotechnol* **2016**, 100, 2035–2046, doi:10.1007/s00253-015-7255-0.

48. Bisson, L.F.; Butzke, C.E. Diagnosis and Rectification of Stuck and Sluggish Fermentations. *Am J Enol Vitic* **2000**, 51, 168–177, doi:10.5344/ajev.2000.51.2.168.

49. Bisson, L.F. Stuck and Sluggish Fermentations. *Am J Enol Vitic* **1999**, 50, 107–119.

50. Lonvaud-Funel, A. Undesirable Compounds and Spoilage Microorganisms in Wine. In *Wine Safety, Consumer Preference, and Human Health* Moreno-Arribas, M., Bartolomé Suáldea, B. Eds. Springer **2016**, 3–26, doi:10.1007/978-3-319-24514-0_1.

51. Bauer, F.F.; Pretorius, I.S. Yeast Stress Response and Fermentation Efficiency: How to Survive the Making of Wine – A Review. *S Afr J Enol Vitic* **2000**, 21, 27–51, doi:10.21548/21-1-3557.

52. Malfeito-Ferreira, M. Wines | Wine Spoilage Yeasts and Bacteria. Encyclopedia of Food Microbiology: Second Edition **2014**, 805–810, doi:10.1016/b978-0-12-384730-0.00390-6.

53. Guillamón, J.M.; Mas, A. Acetic Acid Bacteria. In *Molecular Wine Microbiology* **2011**, 227–255, doi:10.1016/b978-0-12-375021-1.10009-8.

54. Wedral, D.; Shewfelt, R.; Frank, J. The Challenge of *Brettanomyces* in Wine. *LWT - Food Sci Technol* **2010**, 43, 1474–1479, doi:10.1016/j.lwt.2010.06.010.

55. Esteve-Zarzoso, B.; Martínez, M.; Rubires, X.; Yuste-Rojas, M.; Torres, M. Applied Wine Microbiology. In *Molecular Wine Microbiology*. Carrascosa, A.V.; Muñoz, R.; González, R. Ed. Academic Press. **2011**, 341–355, doi:10.1016/b978-0-12-375021-1.10014-1.

56. Compendium of International Methods of Analysis of Wines and Must; 2023 ed.; International Organisation of Vine and Wine: Ensenada (Mexico), **2022**.

57. Delfini, C.; Formica, J. V. Microbiological Monitoring of Must and Wine. In *Wine Microbiology*; CRC Press, **2020**; pp. 289–296.

58. Luong, J.H.T.; Bouvrette, P.; Male, K.B. Developments and Applications of Biosensors in Food Analysis. *Trends Biotechnol* **1997**, 15, 369–377, doi:10.1016/S0167-7799(97)01071-8.

59. Claßen, J.; Aupert, F.; Reardon, K.F.; Solle, D.; Scheper, T. Spectroscopic Sensors for In-Line Bioprocess Monitoring in Research and Pharmaceutical Industrial Application. *Anal Bioanal Chem* **2016**, 409(3), 651–666, doi:10.1007/s00216-016-0068-x.

60. Penner, M.H. Basic Principles of Spectroscopy. **2017**, 79–88, doi:10.1007/978-3-319-45776-5_6.

61. Sun, D.W. Infrared Spectroscopy for Food Quality Analysis and Control; Sun, D.W., Ed.; 1st Ed.; Elsevier, **2009**; ISBN 978-0-12-374136-3.

62. Subramanian, A.; Rodriguez-Saona, L. Fourier Transform Infrared (FTIR) Spectroscopy; 1st Ed.; Elsevier Inc., **2009**; Vol. 1; ISBN 9780123741363.

63. Stuart, B.H. Introduction. In *Infrared Spectroscopy: Fundamentals and Applications*; Stuart, B.H., Ed.; John Wiley & Sons, Ltd: Kent, **2005**; pp. 1–13.

64. Bauer, R.; Nieuwoudt, H.; Bauer, F.F.; Kossmann, J.; Koch, K.R.; Esbensen, K.H. FTIR Spectroscopy for Grape and Wine Analysis. *Anal Chem* **2008**, 80, 1371–1379, doi:10.1021/ac086051c.

65. Craig, A.P.; Franca, A.S.; 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*; Woodhead Publishing, **2015**; pp. 165–194, ISBN 9780857098078.

66. Landgrebe, D.; Haake, C.; Höpfner, T.; Beutel, S.; Hitzmann, B.; Scheper, T.; Rhiel, M.; Reardon, K.F. On-Line Infrared Spectroscopy for Bioprocess Monitoring. *Appl Microbiol Biotechnol* **2010**, 88, 11–22, doi:10.1007/s00253-010-2743-8.

67. 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. In *Measurement, Monitoring, Modelling and Control of Bioprocesses. Advances in Biochemical Engineering/Biotechnology*; Titchener-Hooker, N., Ed.; Springer, Berlin, Heidelberg, **2013**; Vol. 132, pp. 249–280 ISBN 9783642368370.

68. Blanco, M.; Villarroya, I. NIR Spectroscopy: A Rapid-Response Analytical Tool. *TrAC Trend Anal Chem* **2002**, 21, 240–250, doi:10.1016/s0165-9936(02)00404-1.

69. Hashimoto, A.; Kameoka, T. Applications of Infrared Spectroscopy to Biochemical, Food, and Agricultural Processes. *Appl Spectrosc Rev* **2008**, 43, 416–451, doi:10.1080/05704920802108131.

70. Rodriguez-Saona, L.; Ayvaz, H.; Wehling, R.L. Infrared and Raman Spectroscopy. In *Food Analysis*; Nielsen, S.S., Ed.; Springer, Cham, **2017**; pp. 107–127. doi:10.1007/978-3-319-45776-5_8.

71. Griffiths, P.R. Introduction to the Theory and Instrumentation for Vibrational Spectroscopy. In *Handbook of Vibrational Spectroscopy*; Chalmers, J.M., Ed.; Wiley, **2001**; pp. 1–16 ISBN 9780471988472.

72. 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, 303–313, doi:10.1016/j.jfoodeng.2007.12.022.

73. Cynkar, W.; Cozzolino, D.; Dambergs, R.G. The Effect of Sample Storage and Homogenisation Techniques on the Chemical Composition and near Infrared Spectra of White Grapes. *Food Res Int* **2009**, 42, 653–658, doi:10.1016/j.foodres.2009.02.002.

74. Power, A.; Cozzolino, D. Near Infrared for White Wine Analysis. In *White Wine Technology*; Morata, A., Ed.; Elsevier, **2022**; pp. 239–246. doi:10.1016/b978-0-12-823497-6.00009-0

75. Spragg, RA. IR Spectroscopy Sample Preparation Methods. In *Encyclopedia of Specrtoscopy and Spectrometry*; Lindon, J., Ed.; Elsevier Science & Technology, **2016**; pp. 1058–1066. ISBN: 978-0-12-803224-4.

76. 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.

77. Coates, J. A Review of Sampling Methods for Infrared Spectroscopy. *Appl Spectrosc* **1998**, 49–91, doi:10.1016/b978-012764070-9/50005-6.

78. Boyaci, I.H.; Temiz, H.T.; Geniş, H.E.; Acar Soykut, E.; Yazgan, N.N.; Güven, B.; Uysal, R.S.; Bozkurt, A.G.; İlaslan, K.; Torun, O.; Şeker, F.C.D. Dispersive and FT-Raman Spectroscopic Methods in Food Analysis. *RSC Adv* **2015**, 5, 55606–55624, doi:10.1039/c4ra12463d.

79. Schrader, Bernhard.; Bougeard, D. Infrared and Raman Spectroscopy: Methods and Applications; VCH, **1995**; ISBN 978-3-527-61542-1.

80. Abbas, O.; Pissard, A.; Baeten, V. Near-Infrared, Mid-Infrared, and Raman Spectroscopy. In *Chemical Analysis of Food* **2020**, 77–134, doi:10.1016/b978-0-12-813266-1.00003-6.

81. Leary, P.E.; Crocombe, R.A.; Kammrath, B.W. Introduction to Portable Spectroscopy. Portable Spectroscopy and Spectrometry **2021**, 1–13, doi:10.1002/9781119636489.ch1.

82. Crocombe, R.A. Portable Spectroscopy. *Appl Spectrosc* **2018**, 72, 1701–1751, doi:10.1177/0003702818809719.

83. Miller, C.E. Chemometrics in Process Analytical Technology (PAT). Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries: 2nd Ed. **2010**, 353–438, doi:10.1002/9780470689592.ch12.

84. Porep, J.U.; Kammerer, D.R.; Carle, R. On-Line Application of near Infrared (NIR) Spectroscopy in Food Production. *Trends Food Sci Technol* **2015**, 46, 211–230, doi:10.1016/j.tifs.2015.10.002.

85. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.

86. Rinnan, Å.; Norgaard, L.; van den Berg, F.; Thygesen, J.; Bro, R.; Engelsen, S.B. Data Pre-Processing. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Elsevier Ltd, **2009**; pp. 29–50 ISBN 978-0-12-374136-3.

87. Rinnan, Å.; Berg, F. van den; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC - Trend Anal Chem* **2009**, 28, 1201–1222, doi:10.1016/j.trac.2009.07.007.

88. Chuen Lee, L.; Liong, C.-Y.; Aziz Jemain, A. A Contemporary Review on Data Preprocessing (DP) Practice Strategy in ATR-FTIR Spectrum. *Chemometr Intell Lab* **2017**, 163, 64–75, doi:10.1016/j.chemolab.2017.02.008.

89. 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* **2018**, 1058, 9–17, doi:10.1016/j.aca.2018.10.055.

90. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl Spectrosc* **1989**, 43, 772–777, doi:10.1366/0003702894202201.

91. Geladi, P.; MacDougall, D.; Martens, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl Spectrosc* **1985**, 39(3), 491-500.

92. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* **1964**, 36, 1627–1639, doi:10.1021/ac60214a047.

93. Bro, R.; Smilde, A.K. Principal Component Analysis. *Anal Methods* **2014**, 6, 2812–2831, doi:10.1039/c3ay41907j.

94. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C. Breaking with Trends in Pre-Processing? *TRAC Trend Anal Chem* **2013**, 50, 96–106, doi:10.1016/j.trac.2013.04.015.

95. Yun, Y.-H.; Li, H.-D.; Deng, B.-C.; Cao, D.-S. An Overview of Variable Selection Methods in Multivariate Analysis of Near-Infrared Spectra. *TRAC Trend Anal C*hem **2019**, 113, 102–115, doi:10.1016/j.trac.2019.01.018.

96. Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*; Jaumot, J., Bedia, C., Tauler, R., Eds.; Elsevier, **2018**; Vol. 82, pp. 265–299 ISBN 9780444640444.

97. Gerretzen, J.; Szyma Nska, E.; Bart, J.; Davies, A.N.; Van Manen, H.-J.; Van Den Heuvel, E.R.; Jansen, J.J.; Buydens, M.C. 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.

98. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

99. Monakhova, Y.B.; Mushtakova, S.P. Methodology of Chemometric Modeling of Spectrometric Signals in the Analysis of Complex Samples. *J Anal Chem* **2017**, 72, 147–155, doi:10.1134/s1061934816120066.

100. Bartholomew, D.J. Principal Components Analysis. In *International Encyclopedia of Education*, 3rd Ed; Peterson, P., Baker, E., McGaw, B., Eds.; Elsevier, **2010**; pp. 374–377 ISBN 9780080448947.

101. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemometr Intell Lab* **1987**, 2, 37–52, doi:10.1016/0169-7439(87)80084-9.

102. Kucheryavskiy, S. Elements of PCA Model. Accessed: 27th June 2023.

103. Geladi, P. Chemometrics in Spectroscopy. Part 1. Classical Chemometrics. *Spectrochim Acta Part B At Spectrosc* **2003**, 58, 767–782, doi:10.1016/s0584-8547(03)00037-5.

104. Geladi, P.; Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal Chim Acta* **1986**, 185, 1–17, doi:10.1016/0003-2670(86)80028-9.

106. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemometr Intell Lab* **2001**, 58, 109–130, doi:10.1016/s0169-7439(01)00155-1.

107. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA–Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

108. Vitale, R.; Cocchi, M.; Biancolillo, A.; Ruckebusch, C.; Marini, F. Class Modelling by Soft Independent Modelling of Class Analogy: Why, When, How? A Tutorial. *Anal Chim Acta* **2023**, 1270, 341304, doi:10.1016/j.aca.2023.341304.

109. de Juan, A.; Tauler, R. Multivariate Curve Resolution: 50 Years Addressing the Mixture Analysis Problem – A Review. *Anal Chim Acta* **2021**, 1145, 59–78, doi:10.1016/j.aca.2020.10.051.

110. Jackson, J.E.; Mudholkar, G.S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics* **1979**, 21, 341–349, doi:10.1080/00401706.1979.10489779.

111. Diana, G.; Tommasi, C. Cross-Validation Methods in Principal Component Analysis: A Comparison. *Stat Methods Appt* **2002**, 11, 71–82, doi:10.1007/bf02511446/metrics.

112. Fearn, T. Assessing Calibrations: SEP, RPD, RER and R^2. *NIR news* **2002**, 13, 12–13, doi:10.1255/nirn.689.

113. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.M.; McBratney, A. Critical Review of Chemometric Indicators Commonly Used for Assessing the Quality of the Prediction of Soil Attributes by NIR Spectroscopy. *TRAC Trend Anal Chem* **2010**, 29, 1073–1081, doi:10.1016/j.trac.2010.05.006.

114. de Luca, S.; De Filippis, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; 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.

115. Jansen, J.J.; Hoefsloot, H.C.J.; Van Der Greef, J.; Timmerman, M.E.; Westerhuis, J.A.; Smilde, A.K. ASCA: Analysis of Multivariate Data Obtained from an Experimental Design. *J Chemom* **2005**, 19, 469–481, doi:10.1002/cem.952.

116. Ferrer, A. Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift. *Qual Eng* **2013**, 26, 72–91, doi:10.1080/08982112.2013.846093.

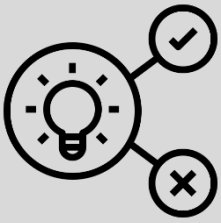
117. Albert, S.; Kinley, R.D. Multivariate Statistical Monitoring of Batch Processes: An Industrial Case Study of Fermentation Supervision. *Trends Biotechnol* **2001**, 19, 53–62, doi:10.1016/s0167-7799(00)01528-6.

118. MacGregor, J.F.; Kourti, T. Statistical Process Control of Multivariate Processes. *Control Eng Pract* **1995**, 3, 403–414, doi:10.1016/0967-0661(95)00014-l.

119. Kourti, T. Process Analytical Technology Beyond Real–Time Analyzers: The Role of Multivariate Analysis. *Crit Rev Anal Chem* **2007**, 36, 257–278, doi:10.1080/10408340600969957.

120. Urtubia, A.; Pérez–correa, J.R.; 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.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García



Chapter 2
Hypothesis and objectives

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Spectroscopic techniques may play a crucial role in the winemaking process by providing valuable insights and enhancing the overall quality of the final product. These techniques, when properly implemented, offer a comprehensive understanding of various stages of winemaking, including grape maturation, alcoholic fermentation, and other critical steps involved in the process.

One of the key advantages of spectroscopic techniques is their ability to provide information about most of the molecules involved and their evolution over time, which enables a rapid response describing the winemaking process. By utilizing spectroscopic analysis decision-makers, such as viticulturists, oenologists, or winemakers, will have access to comprehensive information that would enhance the final product quality, dismiss problems and their possible impacts, and provide the tools to achieve the desired outcomes.

However, one significant challenge that needs to be addressed when using instrumental analytical techniques applied to winemaking, as to every other biological process, is the presence of variability sources that can affect the data obtained. This variability can have many and very different origins such as differences in grape varieties, grape variability in the same field, soil composition, climate conditions, and winemaking practices. If not carefully studied and accounted for, these sources of variability can introduce inconsistencies and diminish the robustness and reliability of results. Therefore, it is crucial to consider and address these sources of variability to ensure the effectiveness of analysis for the control of winemaking process. To consider the process in depth all the experiments in this doctoral thesis were performed at laboratory scale.

Main objective

The general purpose of this PhD Thesis has been to study the vinification process using vibrational spectroscopy coupled with chemometric data treatment. The application of spectroscopy to monitor and control a process requires a prior extensive knowledge of the variability sources to take them into consideration and achieve successful models. This implies the use of spectra generated by mid-infrared, near-infrared and Raman spectroscopy and chemometric techniques. This main objective led to define five sub-objectives.

Specific objectives

- To assess the impact of several positional factors of grapes in the vine within the same field in grape maturity evolution and grape characteristics.

- To evaluate how different stages of grape ripeness and winemaking practices affect the chromatic characteristics of final wines using spectroscopic data collected from grapes and the alcoholic fermentation and maceration processes.

- To determine the performance of Spatially Offset Raman Spectroscopy to monitor wine alcoholic fermentation through a container and predict main oenological parameters.

- To develop chemometric strategies to assess the impact of microbiological contamination on the alcoholic fermentation by acetic acid bacteria (which increases volatile acidity).

- To propose a multivariate statistical tool that enables monitoring wine alcoholic fermentation, detecting sluggish fermentation processes caused by various factors (nitrogen deficit and unwanted temperature changes), and identifying the deviation responsible for the changes in the progression of the process.

Chapter 3
Results

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Section 1

Assessment of Variability Sources in Grape Ripening Parameters by using FTIR and Multivariate Modelling

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Viticulture variability

Grape variability may be due to agronomic parameters that the viticulturist can evaluate and take into account to prevent a negative impact on grape quality and winery profitability[1,2]. Thus, factors such as soil composition, water accumulation or plant diseases, among others, are highly variable and contribute to variations in physicochemical parameters across different field areas[3]. To minimize these effects, the winegrower carries out the necessary soil and vine treatments but, even with all these actions, an exhaustive study of the evolution of grapes in the field is needed to achieve maximum quality.

However, there are other variability sources that cannot be quantified or modified. Climate change has recently emerged as an external factor that significantly and challengingly impacts viticulture[4]. The instability of the climate, characterized by increasingly frequent extreme temperatures and droughts, has caused significant difficulties for viticulturists across the European Union and other wine-producing regions worldwide. This changeable and unpredictable climate makes increasingly challenging for viticulturists to make informed decisions regarding grape cultivation and winemaking practices. The impact of climate change is particularly evident in countries like Spain, where the volume of wine production has decreased 12% in the last five years because drought conditions have become more frequent and severe in recent years[5].

The effects of climate change on grape quality and wine production are becoming more numerous, more evident and more harmful. Thus, rising temperatures accelerate grape ripening, leading to shorter growing seasons and potentially altering the balance between sugars, acidity and phenolic compounds in the grapes (the potential effects of climate change are shown in Figure 1). This can result in unbalanced grapes and wines that lack the desired complexity and character[6]. Furthermore, extreme weather events such as heat waves and torrential rains can cause physical damage to the vineyards, reduce vine yields and increase the risk of diseases and fungi. Drought and lack of rainfall during the critical growth stages can also negatively affect grape development[6].

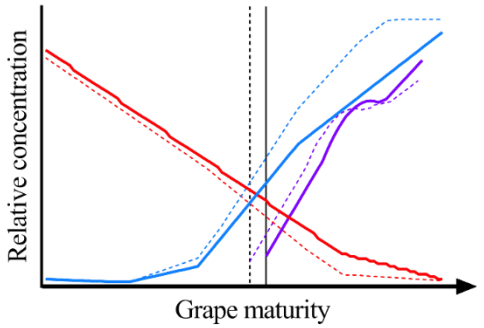


Figure 1. Evolution of sugars (blue), acids (red) and anthocyanins (purple) over grape maturity. Black line represents the veraison. Dashed lines show the tendency of these parameters caused by climate change.

Moreover, the unpredictability of climate change complicates the planning and timing of grape harvest. Traditional viticultural practices rely on field and vine observation throughout several campaigns. It provides the viticulturist with a certain level of predictability regarding grape maturation to determine the optimal harvest date. However, with the shifting climate patterns, the ripening process of grapes becomes more variable, making it difficult to accurately predict the optimum harvest date[6]. Therefore, viticulturists must closely monitor grape physiological indicators and adapt their harvesting decisions accordingly. This flexibility is necessary to ensure that the grapes are harvested at the ideal moment to achieve the desired flavour and balance in the resulting wines[4].

In addition to climate-related challenges, the wine industry also faces problems of overproduction, which may further aggravate the economic difficulties faced by viticulturists. This is because in last years, favourable climatic conditions and optimal grape development may have led to a high yield across vineyards within a Protected Designation of Origin (PDO). This surplus production creates a market imbalance, where the supply exceeds demand. As a result, wineries offer lower prices to viticulturists for their grapes, which often fail to cover the increasing costs associated with growing and harvesting, including rising fuel prices, electricity expenses, and the general cost of living[7].

The economic strain experienced by viticulturists due to overproduction can lead to difficult decisions, as some viticulturists may opt to leave grapes on the vine rather than bearing the costs of harvesting, processing and selling them. This decision can have various consequences, such as the obvious loss of income for the viticulturist but also those

derived from the reduction in vineyard management, which can affect the health and sustainability of the vineyard in the long term[7].

Improvement strategies for the viticulture sector

To mitigate the effects of climate change and overproduction, viticulturists must find strategies to improve their socio-economic circumstances. A good option could be to produce value-added products derived directly from grapes and not from wine, which would provide them with more opportunities and choices to mitigate economic and food losses. Another option could be to take advantage of their extensive knowledge of the field to adopt precision viticulture practices, which could allow them to harvest each grape at its optimum moment in order to achieve the highest quality of the final product with the consequent revaluation of this product.

Diverting grapes towards alternative products

An alternative to redirect wine grapes could be to use them as table grapes for direct consumption as fresh fruit. While wine grapes and table grapes are typically distinct cultivars with different organoleptic properties, there may be cases where wine grapes harvested at a stage not optimal for winemaking can be assimilated as table grapes. This could help alleviate the surpluses of grapes and make them available for consumer consumption, although it would require marketing efforts to educate consumers about this new habit.

Another option could be the production of raisins, which are obtained by drying grapes and produce a flavour rich product. Similar to table grapes, the use of wine grapes in raisin production would not only create an alternative market for surplus grape supply but also would offer consumers a diverse selection of cultivars, resulting in different organoleptic properties for raisin consumption[8].

It should also be taken into account that the different parts of grapes, such as seeds and skins, are rich sources of polyphenols, including compounds like resveratrol and phenolic acids. These bioactive compounds have gained attention for their potential health benefits[9]. As a result, grape juice production using wine grapes or the combination of grapes with other fruits has become an alternative for producing juices with enhanced nutritional and organoleptic properties. Given that the composition of the grape changes

throughout the ripening process, it is not unreasonable to think that, depending on the harvest date, different attributes (both sensory and nutritional) could be obtained, which would allow the production of a wide range of juices[10].

Moreover, grape seeds and skins have shown potential as valuable by-products. These grape parts contain a high concentration of bioactive compounds, making them suitable for various applications[11]. Thus, for example, grape seeds and skins could be used in the production of flours, which can then be used to substitute a portion of wheat or corn flour in the manufacturing of snacks and other food products[12]. This process would not only add value to the viniculture by-products but also would provide an opportunity to develop unique and nutritious food items[9]. Such products, often referred to as "fourth-range products", combine the health benefits of grape-derived compounds while adding more options to the viticulturist to sell the grapes at a fair price.

Precision viticulture

Traditionally, the assessment of grape maturity for winemaking has relied on the collection and determination of some parameters of a representative sample obtained with approximately a hundred of grapes[13]. This approach provides valuable information on critical parameters such as sugar and acidity levels, polyphenolic compounds, and organoleptic properties like seed colour, seed firmness and skin colour. However, the selected sampling method can introduce errors if not executed carefully[14].

A potential problem arises when whole bunches are sampled and then destemmed for analysis. This process may lead to a thinning effect, particularly in vineyards with a small number of bunches. On the other hand, sampling individual berries can also introduce errors due to differences in bunch height within the vine, the different areas of solar irradiation and the presence of compact bunches in which internal grape berries (being difficult to pick) may not be adequately considered. These factors highlight the need for meticulous sampling techniques to ensure representative results[13].

To address the challenges arising from grape variability and to enable more precise monitoring of grape maturity, the implementation of precision viticulture techniques has gained popularity. These techniques encompass remote and proximal sensing techniques, such as infrared spectrometers and hyperspectral imaging cameras, used on plants, berries, soil or combinations thereof. These advanced approaches provide information

about vineyard variability in terms of location and position. By employing these techniques, viticulturists can identify subregions within a vineyard that exhibit unique grape maturity characteristics[15]. In other words, each subregion with distinct maturity profiles can be harvested at the optimal time to maximize grape (and future) wine quality and ensure consistency across the vineyard. Therefore, this approach enables viticulturists to move away from a uniform towards a more tailored and adapted harvest that takes into account the variability in the vineyard[16].

In addition to improving harvest timing, precision viticulture techniques offer other benefits to vineyard management. By identifying spatial variations in grape maturity, viticulturists can target specific areas for interventions such as selective pruning, irrigation adjustments or canopy management. This targeted approach enhances resource allocation, reduces operational costs, and optimizes the overall health and productivity of the vineyard[4,16,17].

The integration of precision viticulture techniques with data analytics and predictive modelling opens up possibilities for real-time monitoring and decision-making. By collecting and analysing data from various sources, such as weather stations, soil sensors, satellites, and grape maturity assessments, viticulturists can gain valuable insights into the factors influencing grape development and adjust management practices accordingly[17].

Study of variability sources

As already stated, exploring alternative pathways for grape utilization not only helps address overproduction, but also provides opportunities for viticulturists to diversify their product offerings and meet changing consumer demands. However, as explained in this chapter, to understand the evolution of grape parameters and to achieve the desired grape properties, it becomes essential to obtain suitable information on the influence of within-field effects on grape characteristics. This requires the use of analytical tools capable of decomposing and quantifying the sources of variability.

One such tool is ASCA, a multivariate extension of ANOVA. ASCA allows for the decomposition of a given dataset, such as the used in the present chapter –parameters or spectra related to grapes– into different matrices that contain specific information related to each factor under investigation[18,19]. By employing a design of experiments, ASCA

enables an objective evaluation of factors while providing quantitative information on their impact. This is why the first study of this Thesis focused on how this chemometric tool could help to unravel the complex interactions between the factors affecting grape maturity, including grape position in the bunch and in the vine. Using ASCA, it was possible to gain a deeper understanding of how these factors contribute to the variability observed in grape maturity and other key parameters[20].

The results of our findings on the influence of berry position within the vine and bunch, as well as its evolution through five weeks around optimal maturity, are gathered in **Paper 2**. Vine and bunches were divided into three parts showing physicochemical and spectroscopic differences that support split harvest to achieve the optimal quality. Furthermore, an overall evolution of the grapes was also observed in the spectra around optimal maturity for wine elaboration. By using these findings, growers can make informed choices to maximize the potential of their grape yield. Additionally, based on this information, alternative grape products can be selected to diversify grape production and optimize the timing of the harvest for specific product outcomes.

Assessment of variability sources in grape ripening parameters by using FTIR and multivariate modelling

Daniel Schorn-García[a], Barbara Giussani[b], María Jesús García-Casas[c], Daniel Rico[c], Ana Belén Martin-Diana[c], Laura Aceña[a], Olga Busto[a], Ricard Boqué[d], Montserrat Mestres[a]

[a]Universitat Rovira i Virgili, Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, 43007 Tarragona, Spain

[b]Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, 22100 Como, Italy

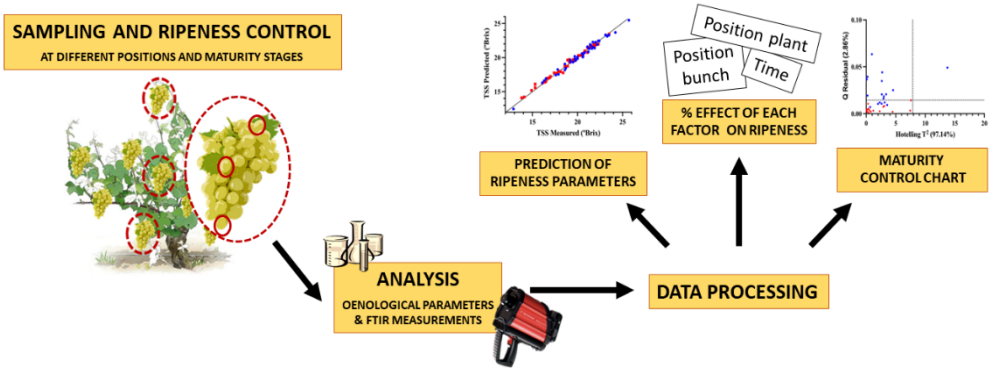
[c]Consejería de Agricultura y Ganadería, Finca de Zamadueñas 47171 Valladolid, Spain

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

Foods 12(5), 962, (2023)

Abstract

The variability in grape ripening is associated with the fact that each grape berry undergoes its own biochemical processes. Traditional viticulture manages this by averaging the physicochemical values of hundreds of grapes to make decisions. However, to obtain accurate results it is necessary to evaluate the different sources of variability, so exhaustive sampling is essential. In this article, the factors "grape maturity over time" and "position of the grape" (both in the grapevine and in the bunch/cluster) were considered and studied by analyzing the grapes with a portable ATR-FTIR instrument and evaluating the spectra obtained with ANOVA–simultaneous component analysis (ASCA). Ripeness over time was the main factor affecting the characteristics of the grapes. Position in the vine and in the bunch (in that order) were also significantly important, and their effect on the grapes evolves over time. In addition, it was also possible to predict basic oenological parameters (TSS and pH with errors of 0.3 °Brix and 0.7, respectively). Finally, a quality control chart was built based on the spectra obtained in the optimal state of ripening, which could be used to decide which grapes are suitable for harvest.



Keywords

portable MIR; variability; ASCA; process monitoring; precision viticulture

Introduction

Wine quality is strongly dependent on the characteristics of the grapes used to elaborate it; therefore, to guarantee their optimal state, a quality control is performed regularly during their ripening process[1]. The common oenological parameters used to follow the evolution of ripening are sugar content and pH, although more detailed information about the ripeness and quality of the grapes can be obtained by determining their phenolic composition, total acidity, texture, and flavor[2]. Setting the optimal ripening point is a key factor in achieving maximum quality in any fruit, but even more so when it comes to non-climacteric fruits such as grapes. This implies that if samples are collected too early, no improvement in any of their quality parameters will be achieved[3], and conversely, if grapes are harvested too late, postharvest disorders are more likely to occur[4].

Another important aspect to consider when working with grapes is the lack of uniformity in the evolution of oenological parameters within the vineyard. This fact has already been shown by different studies in which it has been confirmed that the accumulation of sugars [3,5], anthocyanins[6] or phenolic compounds[7] is not regular in the grapes of the same vineyard[8]. These differences are mainly due to the different viticultural practices — such as soil preparation and tilling, trellising, and pruning of vines or treatments to fight diseases — and to climate and exposure to sunlight, which produces different physiological responses in each plant, in each bunch/cluster and even in each grape[9]. This individual behavior of the grape has led traditional viticulture to so-called precision viticulture, which seeks to know the characteristics of each plant within the vineyard and its evolution over time[10]. However, to obtain this detailed knowledge of the vineyard, it is necessary to carry out a very rigorous and exhaustive sampling followed by a rapid and reliable analysis of the oenological parameters. This entails high costs, both in time and money, so the new trends are focused towards the use of portable (to be able to use them in situ in the field) and rapid analysis techniques together with appropriate statistical treatments that allow choosing those variables that provide significant information to follow the process of interest.

Among the analytical techniques that can be applied on site, vibrational spectroscopy is gaining acceptance because it is fast, robust, and portable[11]. Moreover, as the spectra obtained contain information from almost-all chemical bonds in the sample, when coupled with chemometric data analysis, it provides qualitative and quantitative information on

the composition of the samples under study, often through classification and prediction models[12]. Regarding grape berries, vibrational spectroscopy proved to be suitable for the prediction of several parameters related to technological and phenolic maturities (see [13] and references therein). Additionally, there are chemometric tools that decompose and evaluate the effects of known factors in an experimental design over the multivariate instrumental response such as ANOVA–simultaneous component analysis (ASCA)[14]. ASCA has been applied to spectroscopic data to study the sources of variability during coffee roasting under different conditions[15] or to evaluate chilling injury on the aubergine fruit[16]. Other authors have used ASCA as an exploratory tool to understand which analytical factors directly affect the measurement. Thus, Amigo *et al.* have shown the influence of the measurement area on near-infrared spectra of bread[17] and Borraz-Martínez *et al.* studied the influence of sampling on the discrimination of different varieties of Prunus dulcis leaves[18].

The aim of the present work is threefold. First, assessing the variability sources that affect grape ripeness in terms of harvest time and relative position (both on the plant and on the bunch/cluster) using Fourier transform infrared (FTIR) portable spectroscopy and multivariate analysis. ASCA was used to evaluate the contribution of the different factors and to study their evolution along the maturity process. Secondly, building prediction models of the main parameters related to technological maturity (TSS and pH) from the spectra of the grapes recorded during ripening. Thirdly, pro-posing a strategy based on multivariate statistical process control to determine which grapes have the optimal quality at each moment.

Materials and Methods

Vineyard and Maturity Control

The samples used in this study were obtained from the experimental vineyard of the Faculty of Oenology (Universitat Rovira i Virgili, Spain) located in the Mas dels Frares center (Constantí, Spain) (41°08'44" N 1°12'02" E; Altitude: 60 m; 15 km from the Mediterranean Sea). The climate is characterized by high ambient humidity (60−70%) with hot, dry summers and mild, wet winters. Since the experimental cellar has a record of the evolution of each variety, which includes the harvest date of each vintage, maturity controls usually begin three weeks before the average harvest date. To guarantee the optimal condition of the grapes, these controls involve the measurement of both

technological and phenolic maturity. Technological maturity is determined by measuring the pH and the potential alcoholic strength by the volume (ABV) of the grapes[1], while phenolic maturity is determined by measuring the polyphenol content with the Slinkard and Singleton method[19] and the sensory quality by evaluating the hardness of the skin and the color of the seeds[1]. In addition, taking into account the relevance of the choice of the optimal harvest day for this study, it was decided that the quality of the polyphenols would be evaluated by measuring their antioxidant capacity. Specifically, the oxygen radical absorbance capacity (ORAC) following the method described by Jiménez-Pulido *et al.*[20] was determined.

Samples and Sampling

Ninety grape berries of the variety 'Muscat of Alexandria' were harvested during the entire ripening process. The vineyards are 9 years old and cover 0.65 ha of the Mas dels Frares experimental vineyard, which has a climate that offers the optimal conditions to achieve adequate levels of sugar in the berries of this variety, which are larger than average. In this way, it was guaranteed that the ripening process would develop correctly and that the samples collected would allow obtaining enough juice to carry out the analyses of each berry separately.

A comprehensive sampling was performed to ensure the full monitoring of the ripening process (including overripe samples). Sampling was carried out by collecting 18 grape berries at five different times from 12 August to 15 September (about once a week), as the harvest date was 7 September. This date was considered optimal because the parameters of ripeness, both technological and phenolic, remained practically constant when compared with the values found the previous week and because, although not significant, even a negative trend was sensed in the phenolic parameters (Table 1).

Table 1. Results of the control of ABV (potential alcoholic strength by volume, expressed in alcoholic degrees), pH, TPC (total polyphenolic content, expressed in mg of gallic acid equivalent per 100 mL) and ORAC (oxygen radical absorbance capacity, expressed in µmol of Trollox equivalent per 100 mL). * Values not measured. Different letters indicate significant differences (p-value < 0.05).

Sampling Point	ABV	pH	TPC	ORAC
T1	9.6 ± 1.5 ^{a}	3.15 ± 0.08 ^{ab}	*	*
T2	11.0 ± 1.1 ^{b}	3.10 ± 0.10 ^{a}	10.3 ± 0.6 ^{a}	437 ± 89 ^{a}
T3	12.1 ± 0.4 ^{c}	3.28 ± 0.08 ^{b}	16.8 ± 1.4 ^{c}	620 ± 125 ^{b}
T4	12.4 ± 0.4 ^{c}	3.33 ± 0.13 ^{b}	16.2 ± 1.4 ^{c}	614 ± 125 ^{b}

Grape berries were identified according to the sampling date, the position of the bunch on the plant (top, center, and bottom) and according to the position of the berry within the bunch (top, center, and bottom). Grapes were collected from different plants and bunches to obtain a representative sample of the vineyard. In all cases, a duplicate of the same position relative to plant and bunch, but on different plants, was considered as a biological replicate to ensure maximum variability.

The samples were kept cold until the analysis in portable refrigerators, making sure that the pedicel remained intact to prevent any degradation or evolution.

Determination of Total Soluble Solids and pH in Individual Berries

Once harvested, the samples were immediately taken to the lab, where each individual grape berry was gently crushed with a small manual garlic crusher and the juice obtained was poured directly into 2 mL plastic containers. Analyses of sugar content, expressed as total soluble solids (TSS) were performed at room temperature by using an automatic temperature compensation digital handheld refractometer (HI 96801, Hanna instruments). Before use, the refractometer was calibrated with deionized water and the crystal was thoroughly cleaned with deionized water and wiped dry with cellulose tissues before each new reading. The pH was measured directly into the plastic container with a portable pH meter with a Micro P portable electrode (7+ series portable pH-meter, XS Instruments, Italy). Before analysis, the pH meter was calibrated with two reference standards (pH 7.00 and 4.00).

Mid-Infrared Spectroscopic Analysis

MIR analyses were carried out using a portable 4100 ExoScan FTIR instrument (Agilent, CA, USA), equipped with an interchangeable spherical attenuated total reflectance (ATR) sampling interface consisting of a diamond crystal window and with a diffuse reflectance sampling interface (DRIFT). All samples were analyzed with both interfaces, since the ATR was used to measure the crushed grapes and the DRIFT was used to measure the entire grape berry before crushing.

The spectra were acquired with the Microlab PC software (Agilent, CA, USA) using a methodology previously optimized[21], that is, measuring from 4000 to 650 cm^{-1}, with 32 scans and 8 cm^{-1} resolution.

When using the ATR sampling interface, a drop of each sample (without any pre-treatment) was placed onto the crystal using a plastic Pasteur pipette, ensuring the complete coverage of the ATR crystal, and the spectrum was acquired right afterwards. Each sample was analyzed in triplicate, using three different drops of the crushed grape berry. After each measurement, the crystal was carefully cleaned using deionized water and dried with cellulose tissues. To ensure reliable results, an air background was carried out before each triplicate, i.e., one background before each sample.

For berry measurements using the DRIFT interface, the spectrometer was placed vertically, and the grapes were supported in the sampling gap. Each sample was analyzed in triplicate by placing the grape berry in three different positions (two in the horizontal plane and one opposite the pedicel). In this case, the background was acquired with a 100 Micron Reference (Agilent, CA, USA) after every triplicate.

Data Analysis

Spectral Data Pre-Processing

Spectra were imported with MATLAB (R2021a, 9.10; MathWorks, Natick, MA, USA) to create two datasets (ATR-FTIR and DRIFT), each consisting of 270 spectra (triplicates of ninety samples) and 845 wavelengths. All multivariate analyses were performed using the PLS Toolbox v9.0 (Eigenvector Research Inc., USA). Different pre-processing combinations were tested to mitigate the noise and baseline drifts observed in the raw spectra. The combination that provided the best results for the different models was second order polynomial Savitzky–Golay (SG) smoothing and standard normal variate (SNV). SG is useful for reducing spectral noise, but special attention must be paid to the window size, as severe smoothing could remove useful information from the spectra. SNV was applied to correct spectral light scattering caused by the physical aspects of the sample or equipment characteristics[22,23]. Finally, after spectral pre-processing, data were mean-centered.

Principal Component Analysis (PCA)

PCA was first used to visualize the data and identify outliers by interpreting the score plot and the Hotelling T^2 vs. Q residuals plot (Figure S1). PCA is an exploratory tool that reduces the dimensionality of the data while keeping the maximum information. This

algorithm decomposes the data into a new set of latent variables called principal components, which are linear combinations of the original variables and retain the most information. The projections of the samples onto the new space of principal components (PCs) are called scores, and the angles/projections between the original variables and the PCs are called loadings. Exploring the score and loading plots provides a better understanding of the sources of variability in the spectra and can reveal clusters and trends in the data.

Partial Least Squares (PLS) Regression

PLS regression was applied to predict TSS and pH in the collected samples. The average of the three replicate spectra was used to build the PLS models. An X matrix of spectra (90 samples × 845 wavelengths) was correlated to two y vectors (90 × 1) containing the values for TSS and pH, respectively. For TSS prediction, two approaches were attempted: (1) Using the whole spectrum; and (2) Selecting specific regions based on a previous study [24], which were 967–1175 cm^{-1} and 1483 to 1771 cm^{-1}. To assess model robustness, an external validation was performed. Samples were thus split into calibration and validation sets using the Kennard-Stone[25] and onion[26] algorithms. Two sample splits were used for calibration and validation; half of the data in each set in the first one and 2/3 for calibration and 1/3 for validation in the second one.

The statistical measure root mean square error of prediction (RMSEP) was used (Equation (1)) to evaluate how well the model predicts new samples (not used when building the model):

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_i^{n_t}(y_{t,i} - \hat{y}_{t,i})^2}{n_t}} \qquad \text{Equation 1}$$

$\hat{y}_{t,i}$ is the pH or °Brix predicted by the model, $y_{t,i}$ is the measured value (actual pH or °Brix), and n_t is the number of samples in the test set. RMSEP expresses an average error to be expected in future predictions when the calibration model is applied to unknown samples.

The statistical parameters ratio of performance to deviation (RPD) and range error ratio (RER) were used (Equations (2) and (3)) to evaluate the predictive ability of the models[27]:

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSEP}} \qquad \text{Equation 2}$$

$$\mathrm{RER} = \frac{y_{max} - y_{min}}{\mathrm{RMSEP}} \qquad \text{Equation 3}$$

where SD is the standard deviation of the reference parameters, and y_{max} and y_{min} are the maximum and minimum values of the reference parameters. Both RPD and RER are commonly used to describe whether the obtained prediction models are good enough, with RPD > 2 and RER > 10 indicating good predictive ability.

ANOVA – Simultaneous Component Analysis (ASCA)

Analysis of variance (ANOVA) – simultaneous component analysis (ASCA) is a multivariate exploratory method based on the univariate ANOVA, which decomposes the variability sources affecting the data. ASCA decomposes the overall variation into the main effects and their binary combinations, which are included in a matrix according to a predefined experimental design[28]. The variability sources considered in this study were: (1) Sampling date, which is directly related to the ripening time or maturity status (5 levels); (2) Position of the bunch in the plant (3 levels); (3) Position of the grape within the bunch (3 levels); (4) Spectroscopic replicates (3 levels) and their interactions. The first step of ASCA involves partitioning the centered matrix Xc according to Equation (3):

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = \mathbf{X}_{\mathrm{Ripening}} + \mathbf{X}_{\mathrm{pos.plant}} + \mathbf{X}_{\mathrm{pos.bunch}} + \mathbf{X}_{\mathrm{Ripening\ x\ pos.plant}} + \mathbf{X}_{\mathrm{Ripening\ x\ pos.bunch}} + \mathbf{X}_{\mathrm{pos.plant\ x\ pos.bunch}} + \mathbf{X}_{\mathrm{res}} \qquad \text{Equation 4}$$

where 1 is a vector of ones, $\mathbf{m}^T$ is the average spectrum of the samples, $\mathbf{X}_{\mathrm{Ripening}}$, $\mathbf{X}_{\mathrm{pos.plant}}$, $\mathbf{X}_{\mathrm{pos.bunch}}$ are the matrices of the main factors, $\mathbf{X}_{\mathrm{Ripening\ x\ pos.plant}}$, $\mathbf{X}_{\mathrm{Ripening\ x\ pos.bunch}}$, $\mathbf{X}_{\mathrm{pos.plant\ x\ pos.bunch}}$ are the effect matrices for the binary interactions, and $\mathbf{X}_{\mathrm{res}}$ is the residual matrix collecting all the variability not accounted in the experimental design. Each matrix is centered and contains the mean profiles of the samples corresponding to each factor or interaction level. As an example, if the ripening factor has five levels with 18 observations each, the 18 observations will contain the average profile for the first level of the ripening factor and the same will happen with next level, and so on. The interaction matrix is calculated after the subtraction of the main effect matrices. Afterwards, each matrix is

decomposed using simultaneous component analysis (SCA), which reduces to standard principal component analysis under the constrains of ANOVA[14].

Multivariate Statistical Process Control (MSPC) charts

A PCA-based multivariate statistical process control (MSPC) was applied using the grape spectra of the 4th harvest time, which show the optimal maturity characteristics. Q residuals and Hotelling T^2 values were calculated and plotted in a control chart with 95% confidence limits. The rest of the harvest times were projected onto the PCA (T^2 vs. Q plot) model. The Q statistic indicates how well each sample fits the model and the Hotelling T^2 statistic represents the distance of a given sample to the center of the model[29,30].

Results

Optimization of the Analytical Strategy

Different measurement strategies were tried to obtain suitable instrumental signals. For intact grapes, ATR and DRIFT sampling interfaces were tested (illustrated in Figure S2a,b).

Concerning entire grapes, a problem occurred with the analysis of the most mature samples using ATR. The pressure applied to the sample (needed for a complete contact between the sample and the ATR crystal) caused the skin to break in many experiments. Therefore, we decided not to consider this methodology.

FTIR coupled to the DRIFT interface was tested on entire grape berries even though a low penetration of the radiation in the sample was expected. It is worth recalling that ripeness implies a softening of the grape skin[5], and this information could be captured by DRIFT-IR spectroscopy. However, spectra were noisy (Figure 1a), especially between 2500 and 4000 cm^{-1}.

Individually crushed berries were analyzed with the ATR sampling interface (illustrated in Figure S2c), showing great reproducibility for the instrumental replicates and a good signal-to-noise ratio (Figure 1b). For this reason, the final models were built from the ATR-MIR spectra of crushed grapes.

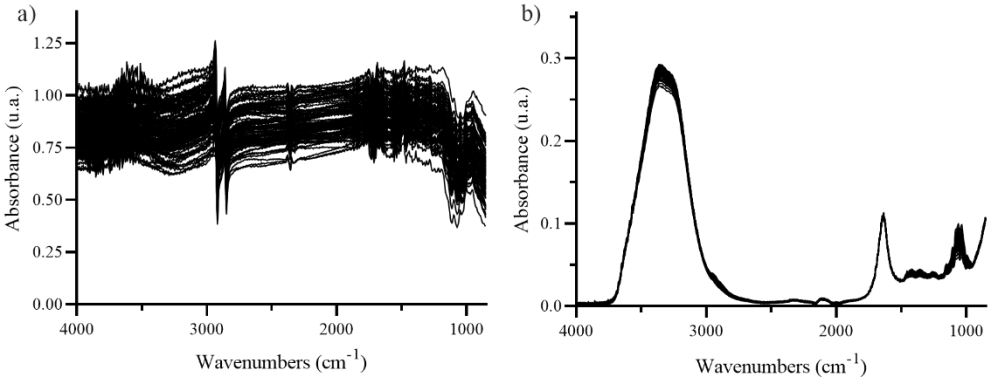


Figure 1. Spectra of 90 grapes using different IR analysis configurations (a) DRIFT for intact berries; (b) ATR-FTIR for crushed berries.

pH and Total Soluble Solids Prediction

The prediction of TSS and pH through multivariate regression required slightly different spectral pre-processing. For the TSS prediction, the pre-processing consisted of SG smoothing with a second-order polynomial and a window width of seven points, followed by SNV. For the pH prediction, the pre-processing consisted of SG smoothing with a second-order polynomial and a window width of fifteen points, followed by SNV. Different validation strategies were applied to assess the robustness of the models[31]. Thus, an external validation set was generated using the Kennard-Stone and onion algorithms and using different ratios between the calibration and validation sets. The PLS model results are shown in Table 2. No outliers were detected.

As shown by the RMSEP and R^2 values, when considering the whole spectra range, TSS was successfully predicted in every model regardless of the methodology to select the samples and the ratio of calibration to validation set sizes. It can be stated that just half of the samples analyzed is enough to obtain good prediction results. Regarding the RMSEP, an error of 0.3 °Brix is quite satisfactory since the TSS range considered goes from 10.1 to 25.7 °Brix, and also because after the potential conversion to potential alcoholic strength, it would mean only an error of ~0.2° (conversion from TSS to ABV according to *Técnicas Analíticas para vinos*[32]. A similar TSS error value has been reported (SECV of 0.20 °Brix) in white grape juices[33]. Additionally, the RPD and RER for TSS prediction were 8.1 and 42.7, respectively. Therefore, the prediction ability of TSS is suitable for assessing grape maturity and is consistent with previous results using ATR-MIR in grapes[34].

Table 2. Prediction results of total soluble solids (TSS) and pH for different validation strategies. RMSEP values for TSS are expressed in Brix. (R^2_{Pred} = determination coefficient of the regression line between predicted and measured values, RS = region selection, LV = number of latent variables of the model), KS = Kennard-Stone algorithm to select the validation set, onion = onion algorithm to select the validation set.

	TSS (°Brix)			**TSS (°Brix)^{RS}**			**pH**		
	RMSEP	**R^2_{Pred}**	**LV**	**RMSEP**	**R^2_{Pred}**	**LV**	**RMSEP**	**R^2_{Pred}**	**LV**
CV random	0.3	0.982	3	0.3	0.981	2	0.07	0.683	9
KS (½ cal–½ val)	0.3	0.984	3	0.3	0.962	2	0.06	0.623	9
KS (⅔ cal–⅓ val)	0.3	0.984	3	0.3	0.979	2	0.06	0.608	9
Onion (½ cal–½ val)	0.3	0.985	3	0.4	0.980	2	0.07	0.687	9
Onion (⅔ cal–⅓ val)	0.3	0.986	3	0.4	0.971	2	0.07	0.591	9

TSS is strongly related to sugar concentration, so the selection of a specific spectroscopic range related to this parameter was likely to produce better results[24]. The results obtained did not show statistically better predictions in terms of RMSEP, R^2, RPD or RER, and even poorer models were obtained when the onion sample selection algorithm was used. However, as for the number of factors, the models built with the selected variables needed two factors instead of three. This could be expected as less spectral information is used for prediction.

In the case of pH prediction, the whole spectra range were used, as previous re-search showed that it provided the best models[24]. This is because variations in pH produce changes in the chemical matrix of the sample, due to changes in bond conformation and matrix properties. As can be seen in Table 2, the prediction error obtained was between 0.06 and 0.07, which is very satisfactory considering the range of the pH values of the samples (between 2.90 and 3.60). This result is comparable to the standard error of cross-validation obtained by Shah *et al*. to predict pH in grape juice samples[33]. Finally, the best models have an RER of 11.5 and an RPD of 2.34, indicating good prediction ability.

The plots of the predicted vs. measured values of the onion external validation using a third of the samples for validation for TSS and pH are depicted in Figure 2.

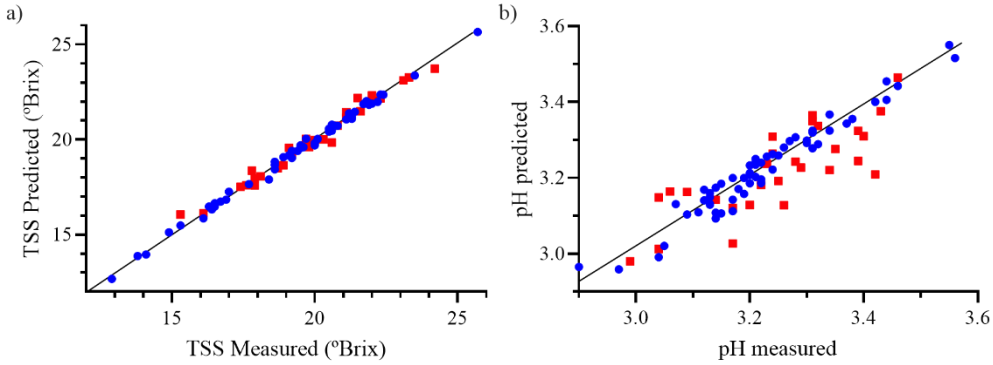


Figure 2. Predicted vs. measured values for the best prediction models of (a) TSS and (b) pH (blue circles for calibration samples, red squares for validation samples).

ANOVA – Simultaneous Component Analysis (ASCA)

An ASCA model was calculated to study the variability sources affecting grapes, using a matrix of 270 samples (90 berries × 3 replicates) and 845 wavelengths (no outliers were detected in the preliminary PCA models).

In the results of an ASCA model, the contribution of each factor in the matrix variability is expressed as a percentage and indicated as % effect. For a given factor, the higher the % effect the more important its contribution. In addition, a permutation test of 10000 iterations was performed to identify significant factors. The significance of the factor is defined by a *p*-value (a *p*-value under 0.05 means the factor is significant). The ASCA results are summarized in Table 3.

First, it should be noted that neither the instrumental replica nor its combination with other factors shows a significant effect, demonstrating that the ATR-FTIR portable device has a high reproducibility in these kinds of measurements. All other factors, maturity (sampling date), position in the plant, position in the bunch, and the interactions between these factors, were statistically significant. Nearly a third of the total variance (28.98%) can be attributed to the sampling time effect. This is consistent with the fact that the evolution of the grape along the ripening process comprises a period of 33 days, in which many physical and chemical changes occur in the samples, including the accumulation of free sugars, cations, amino acids, and phenolic compounds[1].

Table 3. Main effects and their combinations for the ASCA model.

Term	% Effect	*p*-value
Maturity (sampling date)	28.98	0.0001
Position in the plant	5.83	0.0001
Position in the bunch	2.21	0.0268
Instrumental replicate	0.11	0.9116
Maturity × Position in the plant	9.52	0.0001
Maturity × Position in the bunch	5.84	0.0001
Maturity × Instrumental replicate	0.10	1.0000
Position in the plant × Position in the bunch	2.14	0.0224
Position in the plant × Instrumental replicate	0.15	0.9969
Position in the bunch × Instrumental replicate	0.11	0.9987

Figure 3a shows the score values of the maturity submodel for each sample, colored according to different harvest times. Score values are grouped by their sampling time and show an evolution over time. By looking at the first loading of the maturity factor (Figure 3b), this evolution can be assigned to sugars (glucose and fructose), showing a maximum peak at around 1063 cm^{-1} [12,34]. Sugar accumulation and distribution in grapes are major changes in grape samples during ripening[1].

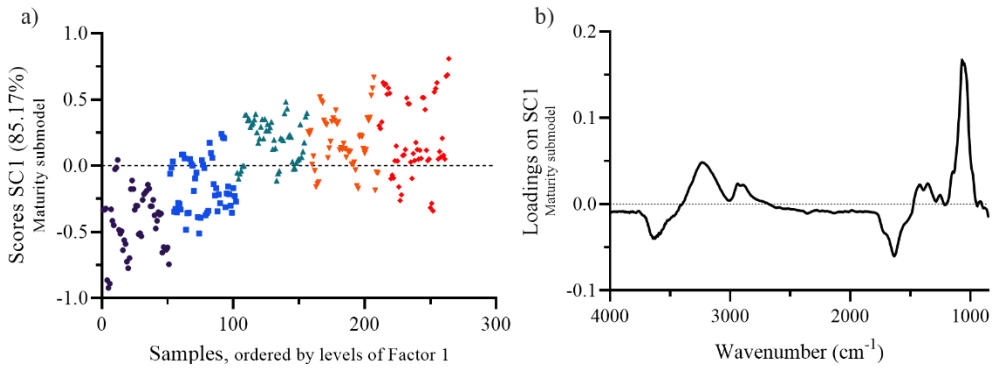


Figure 3. Score (a) and loading (b) plots of the first factor of the maturity factor submodel. Different colors mean that samples belong to different times (Black—Time 1; Blue—Time 2; Green—Time 3; Orange—Time 4; Red—Time 5).

As can be seen in Table 3, the position of the bunch on the plant and the interaction of this factor with the sampling time are the next significant effects in grape spectra after the effect of the sampling time, with an effect of 5.83 and 9.52%, respectively. This result may be due to the different exposure to sunlight depending on the position occupied by the bunch, since there are studies that state that sugar accumulation and final sugar concentration are statistically different depending on the light received by the grape[35]. Additionally, as the leaves are responsible for photosynthesis, which produces the sugars that are accumulated in grapes, the movement of sugars along the plant can affect the different accumulation of sugars and other metabolites depending on the height of the bunch in the vine[36].

In addition, many authors have stated that grape position in the bunch is a factor of variability in terms of sugar concentration and grape size[37,38]. Pagay and Cheng even found different sugar concentrations in three parts of a cluster/bunch. Their results show that the lower part of the bunch has a significantly different sugar concentration compared to the upper part. ASCA results on Table 3 also show that there is a difference between the different parts of the bunch (2.21% effect), and that this difference evolves over time (5.84% for the interaction between grape position in the bunch and sampling time). ASCA results for the interaction between grape position in the bunch and sampling time are in line with those reported in the study by González-Caballero *et al.*, who investigated the impact of ripeness, position of grapes within the bunch and the orientation of the bunch on near-infrared (NIR) spectra[39]. The study revealed differences in the NIR spectra as a function of ripeness, but also due to the position of the grape within the bunch and the orientation of the bunch.

The first loading of all significant factors shows a large signal from 1187 to 937 cm^{-1}, showing that grape ripeness, grape position in the plant, grape position in the bunch, and their interactions are strongly related to the sugar pathway. However, there are other compounds, such as acids, that can also be related to these factors, as there are spectroscopic regions, such as the water band around 3400 cm^{-1} or the fingerprint region between 1800 and 900 cm^{-1}, that also contribute to the models (Figures 3b and S3).

The variance not explained by the ASCA model (residual term) was of 45.01%. This high value could be explained by many abiotic and biotic factors (pluviometry, temperature, plant side, soil, microorganisms, insect action, etc.), as grapes came from an outdoor field[40].

Sub-ASCA (ANOVA – Simultaneous Component Analysis) Models

As explained above, we found that the influence of the factors changed over time, so five individual ASCA models were built to understand how the relative position of the grapes, both in the plant and in the bunch, evolved with time (Table 3). Five subsets of the original matrix, each consisting of 54 spectra and 845 wavenumbers, were used in each ASCA model (results in Table 4). The effect of the instrumental replicates was confirmed not to be significant in any of the models, nor was the interaction of the instrumental replicates with other factors.

Table 4. Results of the ASCA models for each individual harvest period. T1 to T5: number of the different sampling times. * = Significant effect of the factor (p-value < 0.05).

	% Effect				
Sampling Times	**T1**	**T2**	**T3**	**T4**	**T5**
Position in the plant	7.41	40.75*	3.67	36.57*	11.92*
Position in the bunch	22.00*	19.90*	5.52	2.32	5.27
Instrumental replicate	0.23	0.14	0.57	0.28	0.38
Position in the plant × Position in the bunch	46.95*	11.18*	28.98*	4.61	23.97*
Position in the plant × Instrumental replicate	0.44	0.47	1.18	0.23	0.51
Position in the bunch × Instrumental replicate	0.47	0.61	0.95	0.52	0.54
Residual	22.50	26.95	59.14	55.48	57.42

As shown in Table 4, the factors related to grape position evolve over time in a complex manner. The dynamics affecting grape physiology make it difficult to unravel these factors. Grapes, like most fruits, have limited photosynthetic activity, which means that the accumulation of sugars in the berries depends mainly on import from other parts of the plant via the phloem[36]. However, this metabolic input is driven by the individual biochemistry of the grape, as each berry coordinates all the processes necessary in ripening, i.e., sugar accumulation, berry softening, anthocyanin synthesis, metabolism of acids, and accumulation of volatile compounds[5,41]. Therefore, since each berry is

responsible for its own ripening process, there is no synchronization between the berries of the grape, even if they are within the same bunch, which leads to additional variability in the system[42].

The results in Table 4 suggest that there is a distribution between the individual factors (position in the plant and position in the bunch) and the interaction between them. Specifically, when the individual factors are larger, the interaction is lower and vice versa. Furthermore, the effect of the position in the plant is the main effect for three of the sampling times considered (2, 4, and 5), and for the other two sampling times (1 and 3) it is the interaction between the position in the plant and the position in the bunch. This may support the idea that the position in the plant is the most significant effect after sampling time, as seen in Table 3. In addition, temperature differences within the grapes, caused by direct and indirect sun exposure and differences in the top and bottom half of a bunch, significantly affect total soluble solids (TSS) and pH levels[43]. These daily temperature fluctuations, which change throughout ripening, contribute to the overall complexity of the evolution of effects over sampling time, as shown in Table 4[44].

Finally, the residual values indicate that the relative position is the most important factor in the early stages of the ripening process. As time progresses, other factors come into play that cause the residual to increase. This finding suggests that factors other than grape position, such as abiotic and biotic factors, play a more significant role in determining grape physiology as the ripening process progresses.

ANOVA – Simultaneous Component Analysis (ASCA) Model with Reference Parameters

An ASCA model was calculated to study the variability sources affecting grapes in terms of reference parameters. A data matrix of 90 samples and two reference values, pH and TSS, was used for this purpose. Data were autoscaled before applying ASCA. The ASCA results are summarized in Table 5.

ASCA results for pH and TSS show a similar trend to the ASCA results for the spectra, maturity being the most important factor. As shown in Section 3.2, the spectra contain the information of sugars and pH, and this explains the similarity of the results obtained. The effect "position in the plant" could be explained as the difference between the parts of the vine in the metabolites (sugars and acids) imported into the grapes through the phloem[36]. Studies such the one carried out by Doumouya *et al*. have revealed differences between

various parts of the bunch in terms of TSS, pH, TPC, and grape size[45]. The authors emphasize the importance of considering a representative sampling of different parts of the bunch in order to accurately estimate grape ripeness and obtain representative values of the physicochemical properties of the grapes. The small differences between the ASCA models (Tables 3 and 5) could be explained by the fact that the spectra reflect all molecules in the sample, which are not included in the pH and TSS values, such as anthocyanins, which have been shown to be a variability source in white grapes[46]. Despite maturity and position in the plant being the only significant factors in the ASCA model for pH and TSS, every factor shows a % effect similar to that of the ASCA built with the spectra (Table 3, see also Table S1 for the results of TSS and pH for each level in each factor). For that reason, the ASCA model obtained with reference parameters confirms the results obtained with the ASCA model for spectra.

Table 5. Main effects and their combinations for the ASCA model.

Term	**% Effect**	***p*-value**
Maturity (sampling date)	40.49	0.0001
Position in the plant	4.03	0.0300
Position in the bunch	3.10	0.0657
Maturity × Position in the plant	9.07	0.0563
Maturity × Position in the bunch	3.38	0.7729
Position in the plan × Position in the bunch	2.67	0.3663
Residuals	37.25	-

Process Control Charts for Ripening Monitoring

As stated in Sections 3.3 and 3.4, the results of the ASCA models showed that the spatial (position) factors are significant and affect the spectroscopic results. These results have been corroborated using an MSPC approach, based on a control chart built with Q and T^2 statistics[47].

First, a PCA model was built using the mean spectra (three replicates) of the crushed grape samples belonging to Time 4 (matrix of dimensions 18 × 845). Following the assessment of the viticulturist and based on oenological parameters such as ABV, pH, polyphenolic content and antioxidant activity, Time 4 was selected as optimal for the harvest both in terms of technological and polyphenolic maturity. Two PCs were selected

that accounted for around 95% of the variability of the data. The score values of the model were used to calculate the Hotelling T^2 95% confidence limit. To calculate the Q residuals, the residual matrix of the PCA model was used. Individual projections onto the Time-4 PCA model were calculated for the rest of times. In this way, samples with values below the confidence limit are considered similar to those of Time 4, that is, ready to be harvested.

As spatial position factors affect the composition of the grapes, each time had its own variability between the parts of the plant and bunch. It is possible that for times nearer to Time 4, some of the grapes, i.e., some positions, had the same spectral characteristics of Time 4. As expected, all samples from the first two times are above the confidence limit in one or both statistics (Figure S4), thus showing very different characteristics from the samples belonging to the time of harvest (Time 4). For times nearer to the harvest time, some of the samples lie below the confidence limit (Figure 4). Six and five samples lie below the limits for Time 3 and 5, respectively. It is worthwhile to mention that the samples that lie below the limits are not the same in Times 3 and 5 in terms of position in the plant and in the bunch. This would mean a maturity evolution of certain parts of the plant for Time 3 that are similar to Time 4 grapes and certain parts of the plant for Time 5 that are similar to Time 4. However, since there is no uniform behavior in the different positions, we cannot directly decide from the control charts which bunches to harvest already at Time 3 or which ones will be overripe at Time 5.

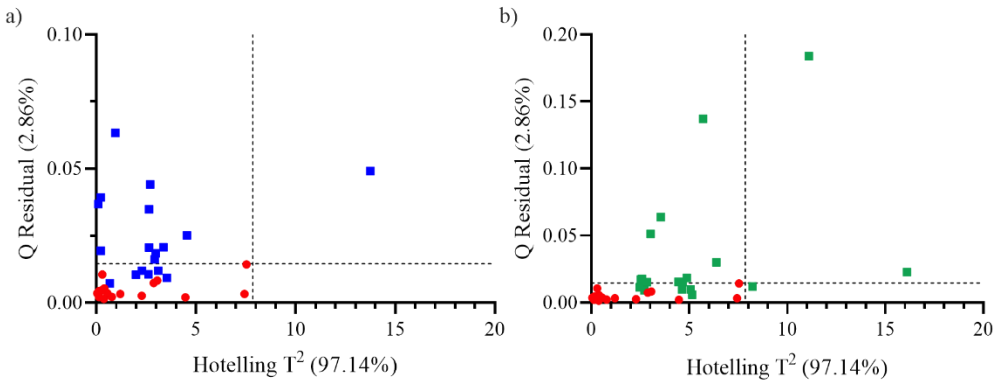


Figure 4. Q residual and Hotelling T^2 for Time 4 (red circles) and the projection in the model of (a) Time 3 (blue squares) and (b) Time 5 (green squares).

From all the presented results we can conclude that the proposed methodology could be used as a quality control chart to select the grapes that are ready to harvest. Thus, this methodology could assess which parts of the plant are ready to be harvested at Time 3 or even wait to be harvested at Time 5; however, it will require more samples in the future and that the study is performed in a specific vineyard to guarantee accurate results.

Conclusions

In this study, a fast and reliable grape ripening control methodology has been developed using a portable ATR-MIR instrument. This methodology has made it possible to predict basic oenological parameters (TSS and pH), which enables the technological maturity of the grapes to be determined. In addition, the different sources of variability in the evolution of grape ripening were also studied using ASCA and it has been objectively shown that the position of the grape, both in the vine and in the bunch, significantly affect ripeness. These results support the idea of splitting the harvest into different days to achieve the best possible quality for each grape. For that reason, we also proposed a control chart to determine which grapes have the optimal characteristics to be harvested at each moment. The results obtained, although preliminary, have shown that this chart could be a useful tool for viticulturists to make the most appropriate decisions about which parts of the vine should be harvested at any given time.

Funding

Grant PID2019-104269RR-C33 funded by MCIN/AEI/ 10.13039/501100011033. This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-García, D.).

Acknowledgments

The authors thanks Ángel García-Pizarro and Raúl Sánchez-Montero for their help in harvesting grapes.

References (Paper 2)

1. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

2. Orlandi, G.; Calvini, R.; Foca, G.; Pigani, L.; Vasile Simone, G.; Ulrici, A. Data Fusion of Electronic Eye and Electronic Tongue Signals to Monitor Grape Ripening. *Talanta* **2019**, 195, 181–189, doi:10.1016/j.talanta.2018.11.046.

3. Shahood, R.; Torregrosa, L.; Savoi, S.; Romieu, C. First Quantitative Assessment of Growth, Sugar Accumulation and Malate Breakdown in a Single Ripening Berry. *OENO One* **2020**, 54, 1077–1092, doi:10.20870/oeno-one.2020.54.4.3787.

4. Kader, A.A. Fruit Maturity, Ripening, and Quality Relationships. *Acta Hortic* **1999**, 485, 203–208, doi:10.17660/actahortic.1999.485.27.

5. Robinson, S.P.; Davies, C. Molecular Biology of Grape Berry Ripening. *Aust J Grape Wine Res* **2000**, 6, 175–188, doi:10.1111/j.1755-0238.2000.tb00177.x.

6. Haselgrove, L.; Botting, D.; van Heeswijck, R.; Høj, P.B.; Dry, P.R.; Ford, C.; Iland, P.G. Canopy Microclimate and Berry Composition: The Effect of Bunch Exposure on the Phenolic Composition of *Vitis vinifera* L. cv. Shiraz Grape Berries. *Aust J Grape Wine Res* **2000**, 6, 141–149, doi:10.1111/j.1755-0238.2000.tb00173.x.

7. Kontoudakis, N.; Esteruelas, M.; Fort, F.; Canals, J.M.; de Freitas, V.; Zamora, F. Influence of the Heterogeneity of Grape Phenolic Maturity on Wine Composition and Quality. *Food Chem* **2011**, 124, 767–774, doi:10.1016/j.foodchem.2010.06.093.

8. Urretavizcaya, I.; Santesteban, L.G.; Tisseyre, B.; Guillaume, S.; Miranda, C.; Royo, J.B. Oenological Significance of Vineyard Management Zones Delineated Using Early Grape Sampling. *Precis Agric* **2014**, 15, 111–129, doi:10.1007/s11119-013-9328-3.

9. Matese, A.; Filippo, S.; Gennaro, D. Technology in Precision Viticulture: A State of the Art Review. *International Journal of Wine Research* **2015**, 7, 69–81, doi:10.2147/ijwr.s69405.

10. Jasse, A.; Berry, A.; Aleixandre-Tudo, J.L.; Poblete-Echeverría, C. Intra-Block Spatial and Temporal Variability of Plant Water Status and Its Effect on Grape and Wine Parameters. *Agric Water Manag* **2021**, 246, 106696, doi:10.1016/j.agwat.2020.106696.

11. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.

12. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

13. dos Santos, C.A.T.; Páscoa, R.N.M.J.; Lopes, J.A. A Review on the Application of Vibrational Spectroscopy in the Wine Industry: From Soil to Bottle. *TRAC Trend Anal Chem* **2017**, 88, 100–118, doi:10.1016/j.trac.2016.12.012.

14. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA Simultaneous Component Analysis: A Tutorial Review. *Anal Chim Acta X* **2020**, 6, 100061, doi:10.1016/j.acax.2020.100061.

15. De Luca, S.; De Filippis, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; 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.

16. Babellahi, F.; Amodio, M.L.; Marini, F.; Chaudhry, M.M.A.; de Chiara, M.L. v; Mastrandrea, L.; Colelli, G. Using Chemometrics to Characterise and Unravel the near Infra-Red Spectral Changes Induced in Aubergine Fruit by Chilling Injury as Influenced by Storage Time and Temperature. Biosyst Eng **2020**, 198, 137–146, doi:10.1016/j.biosystemseng.2020.08.008.

17. Amigo, J.M.; del Olmo, A.; Engelsen, M.M.; Lundkvist, H.; Engelsen, S.B. 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. Food Chem **2019**, 297, 124946, doi:10.1016/j.foodchem.2019.06.013.

18. Borraz-Martínez, S.; Boqué, R.; Simó, J.; Mestre, M.; Gras, A. Development of a Methodology to Analyze Leaves from Prunus Dulcis Varieties Using near Infrared Spectroscopy. Talanta **2019**, 204, 320–328, doi:10.1016/j.talanta.2019.05.105.

19. Slinkard, K.; Singleton, V.L. Total Phenol Analysis: Automation and Comparison with Manual Methods. Am J Enol Vitic **1977**, 28, 49–55.

20. Jiménez-Pulido, I.J.; Rico, D.; Martinez-Villaluenga, C.; Pérez-Jiménez, J.; Luis, D. de; Martín-Diana, A.B. Sprouting and Hydrolysis as Biotechnological Tools for Development of Nutraceutical Ingredients from Oat Grain and Hull. Foods **2022**, 11, 2769, doi:10.3390/foods11182769.

21. Cavaglia, J.; Giussani, B.; Mestres, M.; Puxeu, M.; Busto, O.; Ferré, J.; Boqué, R. Early Detection of Undesirable Deviations in Must Fermentation Using a Portable FTIR-ATR Instrument and Multivariate Analysis. J Chemom **2019**, 33, e3162, doi:10.1002/cem.3162.

22. Rinnan, Å.; Berg, F. van den; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. TRAC Trend Anal Chem **2009**, 28, 1201–1222, doi:10.1016/j.trac.2009.07.007.

23. Rinnan, Å.; Norgaard, L.; van den Berg, F.; Thygesen, J.; Bro, R.; Engelsen, S.B. Data Pre-Processing. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Elsevier Ltd, **2009**; pp. 29–50 ISBN 978-0-12-374136-3.

24. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

25. Kennard, R.W.; Stone, L.A. Technometrics Computer Aided Design of Experiments. *Technometric* **1969**, 11, 137–148.

26. Gallagher, N.B.; O'sullivan, D. Selection of Representative Learning and Test Sets Using the Onion Method. Eigenvector Research Incorporated.

27. Fearn, T. Assessing Calibrations: SEP, RPD, RER and R^2. *NIR news* **2002**, 13, 12–13, doi:10.1255/nirn.689.

28. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

29. Nomikos, P.; MacGregor, J.F. Monitoring Batch Processes Using Multiway Principal Component Analysis. *AIChE Journal* **1994**, 40, 1361–1375, doi:10.1002/aic.690400809.

30. Nomikos, P.; MacGregor, J.F. Multi-Way Partial Least Squares in Monitoring Batch Processes. *Chemometr Intell Lab* **1995**, 30, 97–108, doi:10.1016/0169-7439(95)00043-7.

31. Gorla, G.; Mestres, M.; Boqué, R.; Riu, J.; Spanu, D.; Giussani, B. ATR-MIR Spectroscopy to Predict Commercial Milk Major Components: A Comparison between a Handheld and a Benchtop Instrument. *Chemometr Intell Lab* **2020**, 200, 103995, doi:10.1016/j.chemolab.2020.103995.

32. García Barceló, J. Te□cnicas Anali□ticas Para Vinos; 1a Edición.; GAB: Barcelona, **1990**.

33. 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.

34. Musingarabwi, D.M.; Nieuwoudt, H.H.; Young, P.R.; Eyéghè-Bickong, H.A.; Vivier, M.A. A Rapid Qualitative and Quantitative Evaluation of Grape Berries at Various Stages of Development

Using Fourier-Transform Infrared Spectroscopy and Multivariate Data Analysis. *Food Chem* **2016**, 190, 253–262, doi:10.1016/j.foodchem.2015.05.080.

35. Dokoozlian, N.K.; Kliewer, W.M. Influence of Light on Grape Berry Growth and Composition Varies during Fruit Development. *J Am Soc Hortic Sci* **1996**, 121, 869–874, doi:10.21273/JASHS.121.5.869.

36. Coombe, B.G.; McCarthy, M.G. Dynamics of Grape Berry Growth and Physiology of Ripening. *Aust J Grape Wine Res* **2000**, 6, 131–135, doi:10.1111/j.1755-0238.2000.tb00171.x.

37. Pagay, V.; Cheng, L. Variability in Berry Maturation of Concord and Cabernet Franc in a Cool Climate. *Am J Enol Vitic* **2010**, 61, 61–67, doi: 10.5344/ajev.2010.61.1.61.

38. Tarter, M.E.; Keuter, S.E. Effect of Rachis Position on Size and Maturity of Cabernet Sauvignon Berries. *Am J Enol Vitic* **2005**, 56(1), 86–89, doi: 10.5344/ajev.2005.56.1.86.

39. González-Caballero, V.; Sánchez, M.T.; Fernández-Novales, J.; López, M.I.; Pérez-Marín, D. On-Vine Monitoring of Grape Ripening Using Near-Infrared Spectroscopy. *Food Anal Methods* **2012**, 5, 1377–1385, doi:10.1007/s12161-012-9389-3.

40. Moura, J.C.M.S.; Bonine, C.A.V.; de Oliveira Fernandes Viana, J.; Dornelas, M.C.; Mazzafera, P. Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants. *J Integr Plant Biol* **2010**, 52, 360–376, doi:10.1111/j.1744-7909.2010.00892.x.

41. Armstrong, C.E.J.; Gilmore, A.M.; Boss, P.K.; Pagay, V.; Jeffery, D.W. Machine Learning for Classifying and Predicting Grape Maturity Indices Using Absorbance and Fluorescence Spectra. *Food Chem* **2023**, 403, 134321, doi:10.1016/j.foodchem.2022.134321.

42. Parpinello, G.P.; Nunziatini, G.; Rombolà, A.D.; Gottardi, F.; Versari, A. Relationship between Sensory and NIR Spectroscopy in Consumer Preference of Table Grape (Cv Italia). *Postharvest Biol Technol* **2013**, 83, 47–53, doi:10.1016/j.postharvbio.2013.03.013.

43. Zhang, P.; Barlow, S.; Krstic, M.; Herderich, M.; Fuentes, S.; Howell, K. Within-Vineyard, Within-Vine, and Within-Bunch Variability of the Rotundone Concentration in Berries of *Vitis vinifera* L. cv. Shiraz. *J Agric Food Chem* **2015**, 63(17), 4276–4283, doi:10.1021/acs.jafc.5b00590.

44. Reshef, N.; Fait, A.; Agam, N. Grape Berry Position Affects the Diurnal Dynamics of Its Metabolic Profile. *Plant Cell Environ* **2019**, 42, 1897–1912, doi:10.1111/pce.13522.

45. Doumouya, S.; Lahaye, M.; Maury, C.; Siret, R. Physical and Physiological Heterogeneity within the Grape Bunch: Impact on Mechanical Properties during Maturation. *Am J Enol Vitic* **2014**, 65, 170–178, doi:10.5344/ajev.2014.13062.

46. Agati, G.; Traversi, M.L.; Cerovic, Z.G. Chlorophyll Fluorescence Imaging for the Noninvasive Assessment of Anthocyanins in Whole Grape (*Vitis vinifera* L.) Bunches. *Photochem Photobiol* **2008**, 84, 1431–1434, doi:10.1111/j.1751-1097.2008.00424.x.

47. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Monitoring Wine Fermentation Deviations Using an ATR-MIR Spectrometer and MSPC Charts. *Chemometr Intell Lab* **2020**, 201, 104011, doi:10.1016/j.chemolab.2020.104011.

Supplementary Material

Table S1. Maturity, position in the plant and position in the bunch factors and their levels; the number of samples; TSS and pH values (mean ± standard deviation) of each level of each factor. Different letters mean significant differences between the levels of the factor.

Factor	Level	# Samples	TSS	pH
Maturity	1	18	16.4 ± 2.6 a	3.15 ± 0.08 ab
	2	18	18.8 ± 2.1 b	3.10 ± 0.11 a
	3	18	20.7 ± 1.3 c	3.28 ± 0.11 cd
	4	18	20.6 ± 1.8 bc	3.36 ± 0.13 d
	5	18	21.2 ± 1.9 c	3.25 ± 0.09 bc
Position in the plant	Top	30	18.6 ± 2.9 a	3.19 ± 0.15 a
	Middle	30	19.5 ± 2.8 a	3.24 ± 0.14 a
	Bottom	30	20.2 ± 2.0 a	3.23 ± 0.12 a
Position in the bunch	Top	30	19.6 ± 2.6 a	3.25 ± 0.16 a
	Middle	30	19.4 ± 2.6 a	3.24 ± 0.13 a
	Bottom	30	19.3 ± 2.7 a	3.17 ± 0.12 a

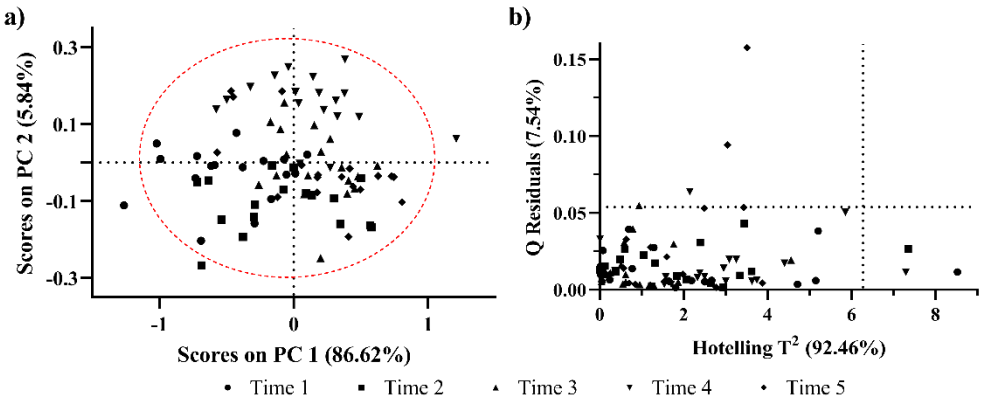


Figure S1. a) PC2 vs PC1 score plot and b) Q residual versus Hotelling T^2 plot of the ATR-FTIR spectra for crushed grapes.

Figure S2. Illustrative photos taken during the analysis of grapes using the different configurations: a) ATR-FTIR for intact grape; b) ATR-FTIR for crushed grape; and c) DRIFT for intact grape.

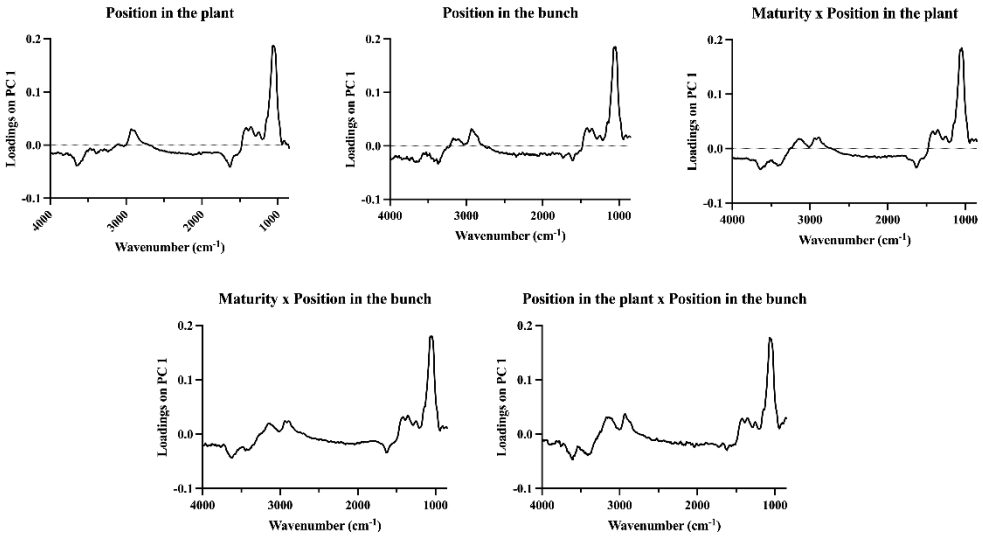


Figure S3. Plots of the first factor of the position in the plan, position in the bunch, interaction maturity x position in the plant, interaction maturity x position in the bunch and interaction position in the plant x position in the bunch factors submodels.

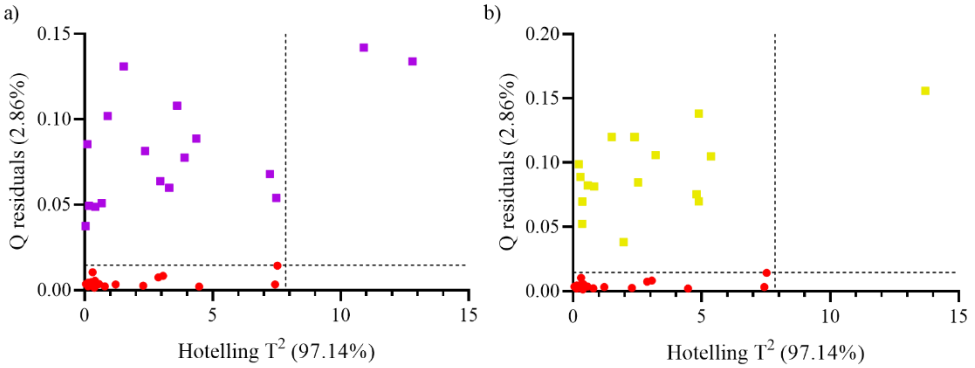


Figure S4. Q residual and Hotelling T^2 for Time 4 (red circles) and the projection in the model of (a) Time 1 (purple squares) and (b) Time 2 (yellow squares).

References (Section 1)

1. Urretavizcaya, I.; Santesteban, L.G.; Tisseyre, B.; Guillaume, S.; Miranda, C.; Royo, J.B. Oenological Significance of Vineyard Management Zones Delineated Using Early Grape Sampling. *Precis Agric* **2014**, 15, 111–129, doi:10.1007/s11119-013-9328-3.

2. Shahood, R.; Torregrosa, L.; Savoi, S.; Romieu, C. First Quantitative Assessment of Growth, Sugar Accumulation and Malate Breakdown in a Single Ripening Berry. *OENO One* **2020**, 54, 1077–1092, doi:10.20870/oeno-one.2020.54.4.3787.

3. Hall, A.; Lamb, D.W.; Holzapfel, B.P.; Louis, J.P. Within-Season Temporal Variation in Correlations between Vineyard Canopy and Winegrape Composition and Yield. *Precis Agric* **2011**, 12, 103–117, doi:10.1007/s11119-010-9159-4.

4. Sun, Q.; Granco, G.; Groves, L.; Voong, J.; Zyl, S. Van Viticultural Manipulation and New Technologies to Address Environmental Challenges Caused by Climate Change. *Climate* **2023**, 11 (4), 83, doi:10.3390/cli11040083.

5. International Organization of Grape and Wine. World Wine Production Outlook. In *OIV First Estimates*; Dijon, **2022**; pp. 1–8.

6. Martínez de Toda, F. Efecto de Las Altas Temperaturas Sobre La Vid, La Uva y El Vino. In *Técnicas vitícolas frente al cambio climático;* Hernández Úbeda, I., Ed.; Ediciones Mundi-Prensa: Santiago de Compostela, **2019**.

7. Catà, J. El Excedente de Cava Desploma El Precio de La Uva Un 30%. *El País* **2019**.

8. Ogletree, K. Father-Son Duo Turns Ruined Grapes into Tasty Aid for Napa Fire Victims. The salt. *What's on your plate* **2018**.

9. Mildner-Szkudlarz, S.; Bajerska, J.; Zawirska-Wojtasiak, R.; Górecka, D. White Grape Pomace as a Source of Dietary Fibre and Polyphenols and Its Effect on Physical and Nutraceutical Characteristics of Wheat Biscuits. *J Sci Food Agric* **2013**, 93, 389–395, doi:10.1002/jsfa.5774.

10. Gülcü, M.; Uslu, N.; Özcan, M.M.; Gökmen, F.; Özcan, M.M.; Banjanin, T.; Gezgin, S.; Dursun, N.; Geçgel, Ü.; Ceylan, D.A. The Investigation of Bioactive Compounds of Wine, Grape Juice and Boiled Grape Juice Wastes. *J Food Process Preserv* **2019**, 43, e13850, doi:10.1111/jfpp.13850.

11. Antonic, B.; Dordevic, D.; Jancikova, S.; Holeckova, D.; Tremlova, B.; Kulawik, P. Effect of Grape Seed Flour on the Antioxidant Profile, Textural and Sensory Properties of Waffles. *Processes* **2021**, 9, 131, doi:10.3390/pr9010131.

12. Giannoutsos, K.; Zalidis, A.P.; Koukoumaki, D.I.; Menexes, G.; Mourtzinos, I.; Sarris, D.; Gkatzionis, K. Production of Functional Crackers Based on Non-Conventional Flours. Study of the Physicochemical and Sensory Properties. *Food Chemistry Advances* **2023**, 2, 100194, doi:10.1016/j.focha.2023.100194.

13. Iland, P.; Bruer, N.; Ewart, A.; Markides, A.; Sitters, J. Monitoring the Winemaking Process from Grapes to Wine: Techniques and Concepts; 2nd Edition.; Patrick Iland Wine Promotions Pty LTD: Adelaide, **2012**.

14. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

15. Mizik, T. How Can Proximal Sensors Help Decision-Making in Grape Production? *Heliyon* **2023**, 9, e16322, doi:10.1016/j.heliyon.2023.E16322.

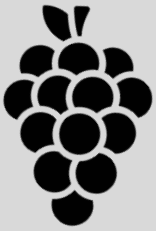
16. Pasquel, D.; Roux, S.; Richetti, J.; Cammarano, D.; Tisseyre, B.; Taylor, J.A. A Review of Methods to Evaluate Crop Model Performance at Multiple and Changing Spatial Scales. *Precis Agric* **2022**, 23, 1489–1513, doi:10.1007/s11119-022-09885-4.

17. Ferro, M.V.; Catania, P. Technologies and Innovative Methods for Precision Viticulture: A Comprehensive Review. *Horticulturae* **2023**, 9, 399, doi:10.3390/horticulturae9030399.

18. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

19. Grassi, S.; Lyndgaard, C.B.; Rasmussen, M.A.; Amigo, J.M. Interval ANOVA Simultaneous Component Analysis (i-ASCA) Applied to Spectroscopic Data to Study the Effect of Fundamental Fermentation Variables in Beer Fermentation Metabolites. *Chemometr Intell Lab* **2017**, 163, 86–93, doi:10.1016/j.chemolab.2017.02.010.

20. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA Simultaneous Component Analysis: A Tutorial Review. *Anal Chim Acta X* **2020**, 6, 100061, doi:10.1016/j.acax.2020.100061.



Section 2

Assessment of the impact of grape maturity and oenological practices on colour extraction and final wine characteristics using infrared spectroscopy

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Red grape cultivars

Phenolic compounds are secondary metabolites found in grapevines, and their biosynthesis is believed to be associated with specific plant needs, such as protection against UV radiation. The defining feature of phenolic compounds, as their name suggests, is the presence of at least one phenol group. In other words, all phenolic compounds contain at least one benzene ring attached to at least one hydroxyl group. From a structural point of view, phenolic compounds can be classified into two main categories: flavonoids and non-flavonoids. Flavonoids are characterized by a shared skeleton consisting of two aromatic rings connected by a pyrene-type heterocycle, and the main compounds are flavonols, anthocyanins and flavan-3-ols, which gives rise to tannins. Non-flavonoids are phenolic acids, such as benzoic and cinnamic acids[1].

In terms of grape maturity, the concept of "technological maturity" is commonly used, which refers to the evolution of sugars and acids over time to achieve the desired concentrations. However, this concept does not consider the grape's composition of secondary metabolites, including aromas and phenolic compounds, which are crucial elements from an oenological perspective. When producing quality red wines, relying solely on technological maturity is insufficient because the optimal expression of aromas and phenolic compounds may not coincide with the moment of optimal technological maturity. For this reason, the term "phenolic maturity" has been coined, which refers to the stage in which the accumulation of polyphenols of interest, synthesized as secondary metabolites, reaches its optimum level giving rise to the best chromatic characteristics, among others, for the wine[2].

Red winemaking variability

In the preceding chapter, we explored the sources of variability in a white grape cultivar. However, when it comes to red cultivars, the sources of variability are even more. The more complex phenolic composition of different red grape cultivars, together with the maturity stages at harvest time introduce a significant level of variability in the winemaking process.

Additionally, red winemaking differs substantially from white winemaking due to the extended contact of grape skins and stems with the fermenting must and wine. This prolonged contact allows for the extraction of phenolic compounds, which provide specific

characteristics of colour, flavour and texture to the red wines. There are diverse methods available for phenolic extraction during the winemaking process, such as varying the extent and the kind of contact between the liquid and solid grape parts or the use of additives to facilitate extraction. Each of these treatments significantly but differently impact the physicochemical and sensory attributes of the resulting wine. This is why these factors greatly contribute to the wide range of red wine styles with different organoleptic properties[2].

Considering the great variability in red grape cultivars, the specific and different maturity of grapes, and the numerous possibilities for phenolic extraction, winemakers face a complex task of managing and optimizing these variables to produce wines of high quality. In addition, it must be taken into account that these variables are not independent of each other, but that there is a clear interaction between grape selection, vineyard practices, winemaking techniques, and the management of phenolic extraction.

Therefore, although it is not an easy task, the final characteristics of the wine related to the phenolic composition can be modulated as long as the expert oenologist has the appropriate viticultural and winemaking tools. These variations in the winemaking process could be addressed to satisfy the different preferences of consumers regarding the phenolic profile of wines[5].

Study of variability sources

Building on the ideas of the previous chapter, where the sources of within-field grape variability were explored using a portable MIR, this chapter focuses on the study of variability sources affecting the chromatic and organoleptic characteristics of the final wine using infrared spectroscopy.

To acquire a comprehensive understanding of the variability sources, both mid-infrared and near-infrared spectroscopy techniques were employed. The spectra obtained from these techniques throughout the winemaking process are used to decompose and quantify the studied sources of variability through experimental design.

The findings of this chapter, encompassing both viticultural influences and oenological practices, are consolidated in **Paper 3**. This study objectively demonstrates that the characteristics of the raw material, the grape itself, have a significant importance not only

on the final characteristics of the resulting wine but also on the evolution of the extraction of polyphenolic compounds. Regarding oenological practices to increase the extraction of the studied compounds, two of them were tested: temperature increase and addition of enzymes. The results showed that oenological practices affect the winemaking process and the final product but these do not take over the inherent characteristics of the grapes. This emphasizes the crucial role of meticulous viticultural management in wine production because, although the oenological tools offer a wide range of possibilities to meet the desired chromatic and sensory attributes in the final product, these will only be achieved if the grapes have the characteristics that allow it. Therefore, these oenological tools enable winemakers to intervene and obtain the desired outcomes only to a certain extent guarantee consistency in the sensory profile of the final product. Thus, they can refine their decision-making processes, optimize viticultural practices, and enhance the quality, distinctiveness and uniqueness of their wines. Finally, we can conclude that this research provides valuable insights into the interplay between viticulture and oenology.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Assessment of the impact of grape maturity and oenological practices on colour extraction and final wine characteristics using infrared spectroscopy

Daniel Schorn–García[a], Kiera Lambrecht[b], Barbara Giussani[c], Ricard Boqué[a], José Luis Aleixandre–Tudó[b,d], Montserrat Mestres[a]

[a]Universitat Rovira i Virgili, Chemometrics and Sensorics for analytical solutions (ChemoSens) group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, 43007 Tarragona, Spain

[b]Department Viticulture and Oenology, South African Grape and Wine Research Institute (SAGWRI), University of Stellenbosch, Stellenbosch 7602, South Africa

[c]Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, 22100 Como, Italy

[d]Insituto de Ingenieria de Alimentos para el Desarrollo (IIAD), Universitat Politècnica de València, Valencia 46022, Spain

Manuscript in preparation

Abstract

The production of red wine involves complex stages and is influenced by various factors, such as the maturity of the grapes, environmental factors and winemaking techniques. Variability in phenolic composition, which plays a crucial role in wine quality due to grape cultivars and winemaking practices, challenges consistent wine production and quality control. Traditional analytical techniques for phenolic analysis are time-consuming and labour-intensive, so alternatives such as mid-infrared (MIR) and near-infrared (NIR) spectroscopy are needed. Understanding the sources of variability, both viticultural and oenological, that affect the phenolic composition and extraction process, facilitates obtaining the desired wine characteristics. Different fermentations were performed using Shiraz, Merlot and Cabernet Sauvignon at different ripening stages and using two oenological practices to increase the extraction of phenolic compounds: addition of enzymes and temperature increase. Decomposition and quantification of variability sources was performed using ANOVA-Simultaneous Component Analysis. Grape cultivar and maturity state were found to be the factors that most influence the final composition of wine, mainly related to differences in the cultivar chemical composition and ethanol production. Phenolic extraction techniques have also a moderate impact on wine characteristics, but they cannot avoid the effect of poor quality raw material.

Keywords
Mid infrared spectroscopy MIR; Near infrared spectroscopy (NIR); variability; ASCA; process monitoring; red wine production

Introduction

Red winemaking involves several complex stages, such as grape ripening, alcoholic fermentation, maceration and aging. They are subject to many factors that can affect their course, as well as the characteristics of final wine obtained[1]. Phenolic compounds, among the many components of wine, make a key contribution to its taste, colour and overall quality. These compounds include a diverse group of secondary metabolites synthesized by grapevines and extracted during winemaking[2]. They can be broadly classified into two main categories: flavonoids and non-flavonoids. Flavonoids, including anthocyanins, flavonols and flavan-3-ols. They influence the colour, astringency, and antioxidant properties of wine. Non-flavonoid phenolics, such as hydroxycinnamic acids and stilbenes, play important roles in wine flavour, stability and health-related attributes[3].

Grape varieties inherently possess distinct phenolic profiles due to the accumulation and evolution of specific phenolic compounds in different grape parts during fruit ripening. Environmental factors, such as climate[4], soil type[5], and viticultural practices[6], further differentiate the phenolic composition of grapes. The interaction between these factors leads to unique phenolic profiles with variations across different wine regions and vintages, and even in the same field[7]. In addition to grape variability, winemaking techniques and processes also significantly influence phenolic composition of the final product. Among other oenological practices[8], must maceration with skins[9], modification of fermentation temperature[10], the use of oak barrels[11] or even microbiological activity during fermentation and aging[12] modulate the extraction, transformation, and stability of phenolic compounds, resulting in diverse phenolic profiles among wines produced even from the same grape cultivar. Therefore, although the expert oenologist faces a complex and highly variable raw material, which challenges for consistent wine production, it must be taken into account that all the available viticultural and oenological tools allow obtaining a great variety of wines with different characteristics that could meet the demand from different consumers[13].

From these premises, there is an increasingly need for robust analytical techniques to understand and monitor the sources of variability in wine production. In the literature there are several methods that use the high-performance liquid chromatography or UV-visible spectrophotometry for the analysis of the phenolic composition of grapes and wines. However, these techniques are time-consuming, labour-intensive, and often require

sample preparation, limiting their applicability for real-time monitoring in a production environment[14].

Spectroscopic techniques, including mid-infrared (MIR) and near-infrared (NIR) spectroscopy, are emerging as alternatives for rapid and non-destructive analysis of phenolic compounds in wine[15]. These techniques provide information that allows the simultaneous quantification and identification of multiple phenolic compounds, which facilitates comprehensive characterization of wine samples. This data interpretation requires suitable chemometric modelling and data analysis that allow the development of robust prediction models based on spectroscopic data for accurate monitoring and prediction of phenolic composition during winemaking[16,17].

Optimizing multivariate approaches to process monitoring and control requires a deep understanding of the variability sources from both a qualitative and quantitative point of view. Thus, this research has a twofold aim: Firstly, to quantify the impact that grape cultivar, its maturity state and the oenological practice applied have on the chromatic characteristics and the phenolic composition of the final wine. Secondly, to monitor the processes of maceration of the must with the skins and of the alcoholic fermentation as a whole by means of Fourier transformed infrared (FTIR) and near infrared (NIR) spectroscopy in order to evaluate and quantify the sources of variability that affect both processes.

Materials and Methods

Grapes and must

Grapes from three cultivars (Shiraz, Cabernet Sauvignon and Merlot) were supplied from "Kleine Liebe" estate, Stellenbosch, South Africa (33°51'06.5"S 18°51'15.3"E, Altitude: 200 m; 34 km from the Atlantic Ocean). For each cultivar, grapes were harvested at two different maturity stages of 2023 vintage: the first sampling was taken approximately two weeks before harvest (pre-ripeness) and the second was collected at the time that the oenologist considered optimal for harvest (optimal ripeness). Each sampling consisted of randomly collecting 240 kg of grapes that were placed in 12 crates of 20 kg of capacity and stored at 4 ºC for two days before processing. To obtain the grape paste to be fermented, grapes from the 12 crates were mixed in a plastic bin to ensure homogeneity and then, grapes were mechanically destemmed and crushed. The crushed grape paste obtained was

divided into 9 parts (triplicates of the 3 types of fermentation explained in the next section) of 20 kg each that were placed in 9 plastic-grade buckets of 25 L, respectively, and added with a solution of potassium metabisulfite to reach a final SO_2 concentration of 30 ppm.

Fermentation management

To start the alcoholic fermentation process, each container was inoculated with $3 \cdot 10^6$ $CFU \cdot mL^{-1}$ of *Saccharomyces cerevisiae* Lalvin ICV D21® (Lallemand, Montreal, QC, Canada), following the instructions of the manufacturer for rehydration of the dry yeast. From each cultivar and maturity state, three types of fermentation were conducted: The first one was kept at 20 ºC during the whole vinification process, and the cap (solid parts of the grape that remain on top of the must) was punched down twice a day; The second one was carried out at 28 ºC with two cap punches per day; The third one was performed at 28 ºC with four cap punches per day and by adding specific enzymes to enhance colour extraction (Lafase® HE Grand CRU Vin Rouge, Laffort, Bordeaux, France) to a final concentration of 1 g per 20 kg of crushed grape paste, meaning per fermentation bucket.

Each fermentation was performed in triplicate, so 54 fermentations (three cultivars, two maturity states, three oenological practices and three biological replicates) were daily followed by measuring density and temperature. When the alcoholic fermentations finished (between the 5th and the 6th day) the subsequent maceration was extended until the 14th day. During this period, 200 mL of each fermentation were collected daily and frozen until analysis.

Spectrophotometric analysis

Each of the different UV-Visible spectrophotometric measurements of the samples were carried out using a Multiskan GO Microplate Spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The parameters determined were: colour density, total anthocyanin content and total phenolic index, SO_2-resistant pigments and tannin content.

Colour density

Fifty microliters of the sample to be analysed were placed in a well of the 96-well microplate the absorbances were measured at 420, 520 and 620 nm. The colour density

was calculated as the sum of the values of the different absorbances (CD = Abs_{420} + Abs_{520} + Abs_{620}). Absorbances were corrected to express the results for a pathlength of 1 cm.

Total Anthocyanin Content and Total Phenolic Index

Total anthocyanin content and total phenolic content were analysed using the method described by P. Iland[18]. 40 microliters of sample were diluted to a final volume of 2 mL with HCl 1M and left in the dark for one hour. After that, 200 microliters were pipetted into a well of a 96-well microplate and the absorbance was measured at 280 and 520 nm. The total phenolic index was directly obtained by applying the dilution factor to the absorbance at 280 nm. The total anthocyanin content, expressed as malvidin-3-glucoside equivalent, was calculated using the following equation:

$$\text{Antocyanin}\ (\text{mg}\cdot\text{L}^{-1}) = \frac{\text{Abs}_{520\text{nm}}\cdot\text{MW}\cdot\text{DF}}{\varepsilon\cdot\text{b}} \qquad \text{Equation 1}$$

where MW and ε are the molecular weight and the extinction coefficient of malvidin-3-glucoside, respectively, DF is the dilution factor and b is the path of the measurement, 1 cm.

SO_2-resistant pigments

SO_2-resistant pigments, which are mainly anthocyanins combined with tannins, were analysed using the modification proposed by Mercurio *et al*[19]. 200 microliters of sample were diluted ten times using a synthetic wine (0.5% tartaric acid (w/v), 12% ethanol (v/v) adjusted to a pH of 3.4 using NaOH) containing 0.375% (w/v) of $Na_2S_2O_5$ and then left in the dark for one hour. After that, 200 microliters were pipetted into a well of the 96-well microplates and the absorbance at 520 nm was measured. The content of SO_2-resistant pigments, expressed as malvidin-3-glucoside equivalent, was calculated using equation 1.

Tannin Content

Tannin concentration was determined using the high throughput[17] method adapted from Mercurio *et al*[19]. This method measures the absorbance difference between the sample (control) and the same sample after precipitation with methylcellulose at 280 nm. 600 microliters of methylcellulose were added to 50 microliters of sample and, after 3 minutes, 400 microliters of $(NH_4)_2SO_4$ and 950 microliters of water were added and left at room temperature for 10 minutes. In the control sample, methylcellulose solution was

substituted by water. Then both samples (treatment and control) were centrifuged at 10000 rpm for 5 minutes and 200 microliters were pipetted into a well of the 96-well microplates and the absorbance at 280 nm was recorded. The absorbance difference was interpolated into a calibration curve using epicatechin (E1753, Merck, Darmstadt, Germany) as standard. The tannin concentration is expressed as mg of epicatechin equivalent per litre.

Mid-Infrared Spectroscopic Analysis

MIR analyses were carried out using an Alpha-P FTIR instrument (Brucker Optics, Ettlingen, Germany), equipped with an attenuated total reflectance (ATR) sampling interface consisting of a single bounce diamond crystal window. The sampling device was kept at 30 ºC in every analysis to ensure reproducibility between samples. The spectra were acquired with the OPUS Wine Wizard (OPUS v7.0 for Microsoft, Bruker Optics, Ettlingen, Germany) using a methodology previously optimized[17], that is, measuring from 4000 to 400 cm^{-1}, with 128 scans and 4 cm^{-1} resolution. Due to the great number of samples, several days of eight to ten hours of work were needed for analysis, and to ensure reliable results, a water background was carried out every two hours and samples were previously centrifuged.

Near-Infrared Spectroscopic Analysis

NIR analyses were carried out using a Multi-purpose analyser FT-NIR instrument (Brucker Optics, Ettlingen, Germany), in transmission mode using a 1 mm quartz cuvette. The spectra were acquired with the OPUS Wine Wizard using a methodology previously optimized[17], that is, measuring from 12500 to 4000 cm^{-1}, with 64 scans and 1 cm^{-1} resolution. As per MIR spectra collection to ensure reliable results, a water background was carried out every two hours and samples were previously centrifuged.

Multivariate Analysis

Principal Component Analysis (PCA)

PCA was first used to visualize the data and identify potential outliers, by interpreting the score and influence (Hotelling T^2 vs. Q residuals) plots. PCA is an exploratory tool that reduces the dimensionality of the data while keeping the maximum information.

Additionally, exploring the score and loading plots provides a better understanding of the sources of variability in the spectra and can reveal clusters and trends in the data.

ANOVA– Simultaneous Component Analysis (ASCA)

Analysis of variance (ANOVA)–simultaneous component analysis (ASCA) is a multivariate exploratory method based on the univariate ANOVA, which decomposes the variability sources affecting the data. ASCA decomposes the overall variation into the main effects and their binary combinations, which are included in a matrix according to a predefined experimental design[20]. The variability sources considered in this study were: (1) Grape cultivar (3 levels); (2) Maturity status (2 levels); (3) Fermentation day (15 levels); (4) Oenological practices (3 levels) and their interactions. The first step of ASCA involves partitioning the centred matrix $\mathbf{X}$c according to Equation (3):

$$\begin{aligned}\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = {} & \mathbf{X}_{\text{grape cultivar}} + \mathbf{X}_{\text{grape maturity}} + \mathbf{X}_{\text{fermentation}} + \mathbf{X}_{\text{oeno.practices}} \\ & + \mathbf{X}_{\text{cultivar x maturity}} + \mathbf{X}_{\text{cultivar x fermentation}} + \mathbf{X}_{\text{cultivar x oeno.practices}} \\ & + \mathbf{X}_{\text{maturity x fermentation}} + \mathbf{X}_{\text{maturity x oeno.practices}} \\ & + \mathbf{X}_{\text{fermentation x oeno.practices}} + \mathbf{X}_{\text{res}}\end{aligned} \qquad \text{(Equation 4)}$$

where $\mathbf{1}$ is a vector of ones, $\mathbf{m}^T$ is the average spectrum of the samples, $\mathbf{X}_{\text{grape cultivar}}$, $\mathbf{X}_{\text{grape maturity}}$, $\mathbf{X}_{\text{fermentation}}$, $\mathbf{X}_{\text{oeno.practices}}$ are the matrices of the main factors, $\mathbf{X}_{\text{cultivar x maturity}}$, $\mathbf{X}_{\text{cultivar x fermentation}}$, $\mathbf{X}_{\text{cultivar x oeno.practices}}$, $\mathbf{X}_{\text{maturity x fermentation}}$, $\mathbf{X}_{\text{maturity x fermentation}}$, $\mathbf{X}_{\text{fermentation x oeno.practices}}$ are the effect matrices for the binary interactions, and $\mathbf{X}_{\text{res}}$ is the residual matrix collecting all the variability not accounted in the experimental design. Each matrix is centred and contains the mean profiles of the samples corresponding to each factor or interaction level. As an example, if the grape maturity has two levels with 405 observations each, the 405 observations will contain the average profile for the first level of the ripening factor and the same will happen with next level, and so on. The interaction matrix is calculated after the subtraction of the main effect matrices. Afterwards, each matrix is decomposed using simultaneous component analysis (SCA), which reduces to standard PCA under the constraints of ANOVA[21].

Results

Final wine characterization

A phenolic characterization of each sampling point as well as the final wine was performed in order to understand the evolution of the different parameters related to colour as well as to characterize the final wine obtained. Table 1 shows the results of total anthocyanin content, SO_2 resistant pigments, tannins, colour density and Total Phenolic Index for the final wines split according to the factors considered.

Table 1. Grape cultivar, grape maturity and oenological practices and their levels; anthocyanin content (expressed as mg of malvidin-3-glucoside equivalent per litre), SO_2–resistant pigments (expressed as mg of malvidin-3-glucoside equivalent per litre), tannins (expressed as mg of epicatechin equivalent per litre), colour density and Total Phenolic Index (TPI) values (mean ± standard deviation) of each level of each factor. Different letters mean significant differences between the levels of the specific factor (p-value < 0.05).

Factor	Level	Anthocyanin ($mg \cdot L^{-1}$)	SO_2 resistant pigments	Tannins ($mg \cdot L^{-1}$)	Colour density	TPI
Grape cultivar	Merlot	282 ± 56 a	26.6 ± 9.8	955 ± 190 a	9.3 ± 1.9 a	23.3 ± 5.0 a
	Shiraz	295 ± 56 a	21.6 ± 9.8	874 ± 190 a	8.2 ± 1.9 a	24.5 ± 5.0 a
	C. Sauvignon.	348 ± 58 b	22.1 ± 10.1	1211 ± 195 b	11.4 ± 2.0 b	32.7 ± 5.2 b
Grape maturity	Pre-ripeness	269 ± 56 a	22.1 ± 9.8	940 ± 190 a	7.7 ± 1.9 a	23.9 ± 5.0 a
	Optimal ripeness	348 ± 57 b	24.8 ± 10.0	1087 ± 193 b	11.6 ± 2.0 b	29.8 ± 5.1 b
Oenological practices	20 ºC	308 ± 58	17.7 ± 10.1 a	829 ± 195 a	8.2 ± 2.0 a	22.6 ± 5.2 a
	28 ºC	316 ± 56	26.3 ± 9.8 b	1097 ± 190 b	10.3 ± 1.9 b	29.5 ± 5.0 b
	28 ºC + Enzyme	302 ± 56	26.3 ± 9.8 b	1114 ± 190 b	10.3 ± 1.9 b	28.5 ± 5.0 b

Significant differences were observed in the phenolic composition and colour density of the Cabernet Sauvignon grape cultivar compared to other varieties. Except for SO_2–resistant pigments, all parameters showed significantly higher values for Cabernet sauvignon, while Merlot and Shiraz showed similar values for all parameters.

Furthermore, as expected, grape maturity state had a notable impact on both phenolic and colour parameters. This can be attributed to the continuous evolution of phenolic maturity

until harvest, resulting in higher concentrations of various phenolic compounds and higher extractability[7,22]. Another factor contributing to these differences could be the ethanol content, as phenolic compounds tend to be extracted more effectively in high ethanol environments[23].

Finally, it should be noted that both oenological practices studied also have a significant impact on wine characteristics. Interestingly, the highest anthocyanin content and TPI value were obtained for wines fermented at 28 ºC without the use of enzymes. This can be attributed to different factors. First, in the production of wines using the high extraction of polyphenols method (28 ºC and use of enzymes), anthocyanins may undergo degradation, precipitation, or absorption by container surfaces or yeast cell membranes[24,25]. Second, high concentrations of anthocyanins may stimulate the production of other pigments with different maximum wavelengths. In cases where enzymes were used, a slight tendency towards higher tannin values was observed, as the enzymes facilitated the extraction of tannins predominantly found in grape seeds.

There results show that there are multitude of factors, possibilities and interactions that allow obtaining very different products. However, for this study, the importance of these results lies in the fact that, starting from the same raw material, different products can be obtained depending on how the winemaking process has been carried out. Therefore, we can follow these processes and assess which factors have the most influence on each of them.

Spectroscopic evaluation

Figure 1 shows the mean spectra of the two spectroscopic techniques (MIR and NIR, respectively) for the different cultivars in the upper row (Fig 1a-b) and the different maturity states (Fig 1c-d).

MIR spectra for the different cultivars showed a similar profile for Merlot and Shiraz, and slightly higher signal for C. Sauvignon. This is especially remarkable around 1100 cm^{-1}, the spectroscopic band related to C–O[26]. This could imply a higher technological maturity for this cultivar as this signal is usually attributed to sugars contribution[26]. Regarding NIR spectra, due to the more complex interpretability compared to MIR spectra, and accentuated by a difference in the baseline, the possibility to assess spectroscopic differences between cultivars was difficult.

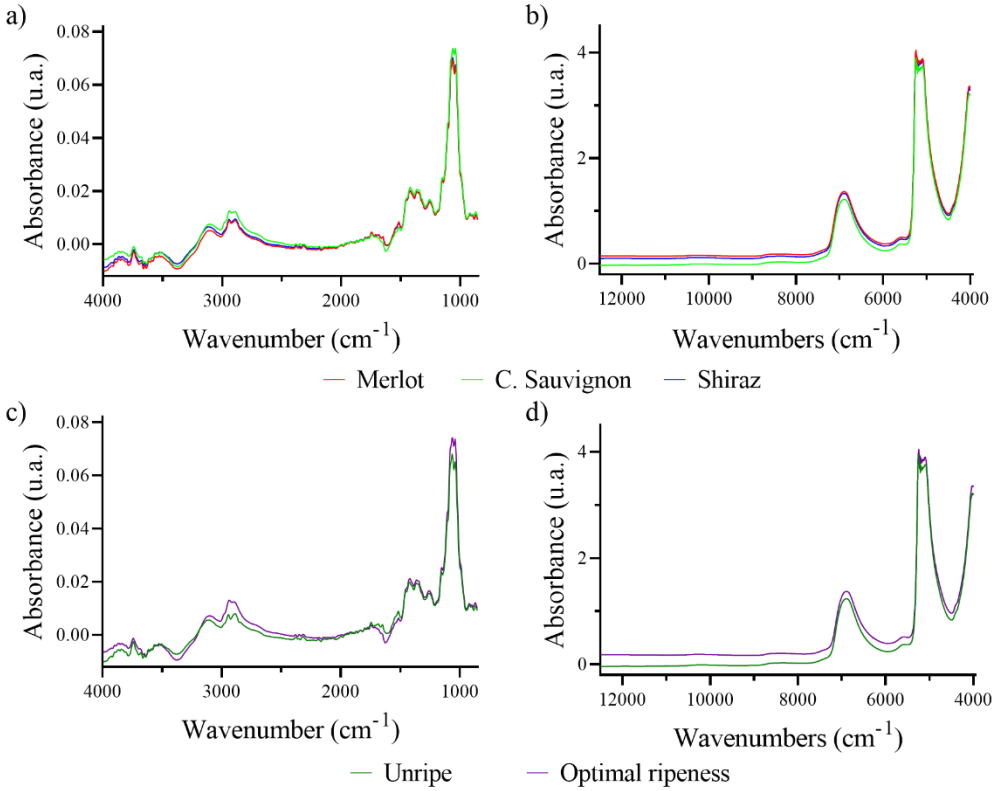


Figura1. Mean spectra of grape must from the different grape cultivars acquired by (a) MIR and (b) NIR, and from the different maturity (c) MIR and (d) NIR.

Regarding maturity states, as expected, optimal ripeness also showed a higher signal in the C–O band which could be related to a higher concentration of sugars[26]. NIR spectra had a baseline difference between maturity states, which is even more pronounced than when comparing varieties, further complicating the possibility to assess spectroscopic differences.

Study of the effects in the extraction process

An ASCA model was built to study the influence of the different factors affecting the winemaking evolution of phenolic composition and colour characteristics. To build the ASCA model a matrix containing 810 observations or samples (no outliers were detected in the preliminary PCA models) and 5 variables (anthocyanin content, SO_2–resistant pigments, tannins, colour density and Total Phenolic Index values) was used. Data was autoescalated prior modelling.

To interpret the ASCA results (Table 2), two things must be kept in mind: the contribution of each factor to the matrix variability, expressed as a percentage (% effect in Table 2) and the significance of the factor defined by a *p*-value. To identify significant factors a permutation test of 10000 permutation was performed. A factor was considered significant when the *p*-value was under 0.05.

Table 2. Main effects and their combinations for the ASCA model.

Term	**% Effect**	***p*-value**
Variety	6.95	0.0001
Maturity	6.69	0.0001
Oenological practice	4.14	0.0001
Winemaking day	37.22	0.0001
Variety x Maturity	7.36	0.0001
Variety x Oenological practice	0.64	0.0001
Variety x Winemaking day	3.93	0.0001
Maturity x Oenological practice	0.22	0.0034
Maturity x Winemaking day	2.25	0.0001
Oenological practice x Winemaking day	3.05	0.0001

First, it is important to highlight that all factors and their binary combinations show significant effects on the phenolic composition and colour evolution process. Among these factors, the most influential (37.22%) is the winemaking day. This is consistent with the fact that alcoholic fermentation, which occurs during the early stages of winemaking, generates ethanol that facilitates the extraction of phenolic compounds. Furthermore, it has been reported that the duration of maceration also has a significant impact on the concentrations of anthocyanins, tannins, and polymeric pigments (SO_2–resistant pigments)[24,25]. However, due to the relatively short maceration time, no major accumulation of this compounds was observed in the time interval considered. The reaction of combination between anthocyanins and tannins requires time to occur[27].

Figure 2 shows the score evolution over time for the winemaking day submodel. As can be seen, the greatest evolution of the scores occurs in the first days of the process. This can be attributed to the increasing concentration of ethanol, particularly in the early stages of

alcoholic fermentation. Additionally, a slight decrease in the scores is observed during the later stages of the process. As maceration time progresses, there is an inverse effect on phenolic compounds, as they begin to polymerize and overextraction causes them to undergo processes of precipitation or absorption by yeast and/or the container[24,25].

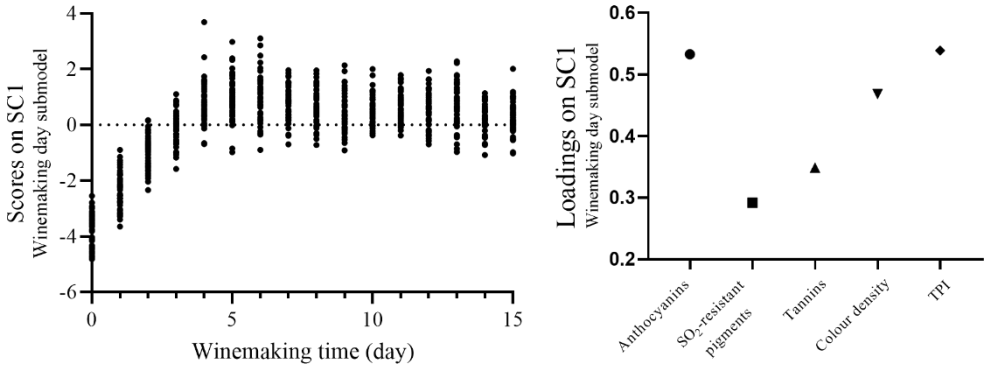


Figure 2. Score plot of the first simultaneous component of the winemaking day submodel over process time.

Optimization of the maceration time is crucial to achieve the best phenolic composition, because phenolic compounds have been related to wine quality[28]. The loadings of the submodel showed that all the variables contribute positively to the model, with the greater contribution of anthocyanins and TPI values.

The next factors having a significant impact on the evolution of phenolic compounds are grape cultivar, maturity and their combination. As time has the greatest impact on the extraction of phenolic compounds, these factors do not represent a high effect. However, their effects underscore the importance of grape characteristics in the winemaking process. The quality of the grapes, including their maturity, plays a critical role in determining the composition of these compounds.

Oenological practices have a lower impact on the evolution of phenolic compounds than grape cultivar and maturity. These findings emphasize the need for careful attention in the vineyard, as winemaking alone may not fully guarantee the desired phenolic composition. Despite that, oenological practices have a significant impact on the final wine, which means that wineries have the ability to carry out the fermentation according to the desired characteristics taking into account grape characteristics.

Finally, the binary combinations of the winemaking day factor with other factors also show a significant contribution. This suggests that each grape cultivar, at different levels of maturity and subject to various vinification styles, will exhibit unique evolution during the winemaking process. This observation aligns with the diverse range of wine styles available in the market, highlighting the complexity and variability inherent to winemaking. However, an exhaustive study of the sources of variability makes it possible to take them into account, control them and make decisions to achieve the desired results.

Spectroscopic study of the factors affecting wine phenolic composition

Two ASCA models were calculated to study the variability sources affecting winemaking, using two data matrices, one matrix of MIR spectra and another one of NIR spectra for every fermentation in every sampling point. Pre-processing techniques consisted in smoothing (Savitsky-Golay 2nd order 15 points) for MIR spectra and detrending (1st order) for NIR spectra. The ASCA results are summarized in Table 3.

Table 3. Main effects and their combinations for the ASCA models with MIR and NIR spectra. *: the effect is statistically significant (p-value < 0.05).

Term	% Effect MIR	% Effect NIR
Variety	0.41*	2.03*
Maturity	1.21*	3.29*
Oenological practice	0.47*	0.26
Winemaking day	59.44*	26.50*
Variety x Maturity	1.13*	3.48*
Variety x Oenological practice	0.11	0.23
Variety x Winemaking day	1.76*	4.68*
Maturity x Oenological practice	0.17	0.20
Maturity x Winemaking day	6.28*	7.75*
Oenological practice x Winemaking day	1.63*	2.99*

A general trend similar to the ASCA results for phenolic compounds is observed for both ASCA models. The factor "winemaking day" is the one that most affects the process, in the case of ASCA for MIR spectra with the greatest effect. This is due to the evolution of other compounds present in the matrix that are present in the spectra. Both spectroscopic

techniques show an evolution of the scores for the winemaking day submodel (Figure 3), which is similar to the one observed for phenolic compounds. By looking at the first loading of the winemaking day factor for both spectroscopic techniques, the evolution of the scores can be assigned to both sugars and phenolic compounds, as the fingerprint region in MIR spectra shows a certain contribution to the model[26].

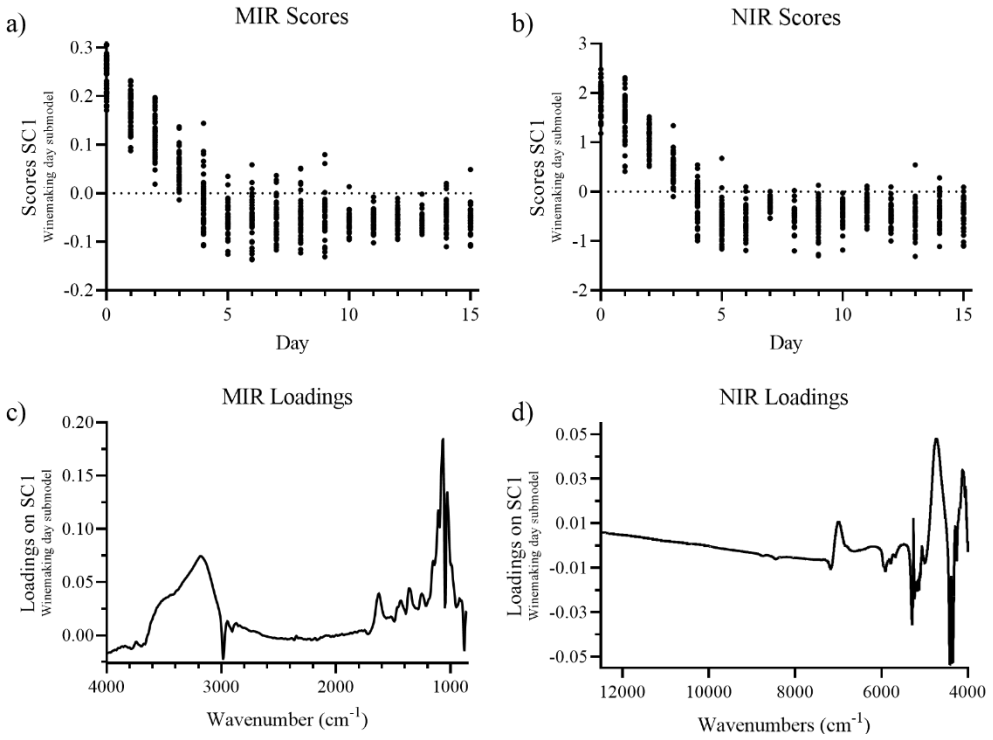


Figure 3. Scores of the first component of the factor "winemaking day" for the ASCA model using (a) MIR spectra and (b) NIR spectra. Loadings plot of the first component of the factor "winemaking day" for (c) MIR spectra and (d) NIR spectra.

For both spectroscopic techniques, similarly to the results for phenolic compounds, the factors "grape variety" and "grape maturity" are the next factors in terms of % effect. Despite of explaining less variability in the ASCA model for NIR spectra, both effects are greater than in MIR. This could be explained because sugars have a greater absorption in MIR that may hinder information of other chemical bonds[29]. The binary interaction of both factors has also a significant effect on the evolution. Additionally, transmittance-mode NIR spectroscopy has been previously reported to underperform in the prediction of phenolic compounds, which could explain the ASCA results obtained[16,17].

The factor "oenological practices" was found to be significant only in the ASCA model for MIR spectra. This factor had the smallest effect within the individual factors in the ASCA model for phenolic compounds, which could explain its lower effect on the spectra due to the overlap with other chemical bonds. For NIR spectra, neither this factor nor its binary combination with the rest of factors provided a significant effect.

Finally, the binary combinations of the factor "winemaking day" with the factors "grape variety" and "maturity", showed the specific evolution of each fermentation for each cultivar. It has previously been reported that grape variability could be detected using spectroscopy within the same field[30,31]. The variance not explained by the ASCA model could be due to many factors such as yeast metabolism, differences between replicates, or instrumental factors such as light scattering or instrumental noise[31,32].

Based on the results presented, it can be concluded that both methodologies are able of helping to understand the sources of variability in red winemaking, except for oenological practices using NIR spectroscopy. Additionally, as the three ASCA models show a similar evolution for the factor "winemaking day", it is thus possible to monitor red winemaking using spectroscopy, as previously reported[16,33], taking into account what is the impact of raw materials and winemaking practices.

Conclusions

In this study, we performed a quantitative assessment of the influence of grape cultivar and grape quality, specifically phenolic maturity, on the characteristics of the final wine. The findings revealed that the raw material, i.e., the grapes and their maturity state, had a more significant impact on the phenolic composition of the wine compared to other factors. Various oenological practices with different extraction abilities were also examined, and although they were found to have a significant impact, their influence was lower than that of grape composition.

Interestingly, it was observed that the practices with highest extraction resulted in lower phenolic concentrations compared to the practices with medium extraction, regardless of grape cultivar and maturity. This suggests that overextraction occurred, leading to reactions, mainly precipitation, which affected the phenolic compounds. Furthermore, the evolution of phenolic compounds during alcoholic fermentation and post-fermentative

maceration was investigated, confirming the role of ethanol and the duration of the maceration in the extraction process.

Although these results are preliminary, they highlight the potential of mid-infrared and near-infrared spectroscopy in studying the variability sources in the production of red wine. The application of these spectroscopic techniques can be potentially extended to the winemaking industry, so they can provide valuable information about the production of red wines. Further research and exploration will be surely developed to fully understand the capabilities of these analytical tools in enhancing winemaking practices and producing wines with increased quality.

Funding

Grant PID2019-104269RR-C33 funded by MCIN/AEI/ 10.13039/501100011033. This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-García, D.).

References (Paper 3)

1. Setford, P.C.; Jeffery, D.W.; Grbin, P.R.; Muhlack, R.A. Factors Affecting Extraction and Evolution of Phenolic Compounds during Red Wine Maceration and the Role of Process Modelling. *Trends Food Sci Technol* **2017**, 69, 106–117, doi:10.1016/j.tifs.2017.09.005.

2. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

3. Cheynier, V.; Dueñas-Paton, M.; Salas, E.; Maury, C.; Souquet, J.M.; Sarni-Manchado, P.; Fulcrand, H. Structure and Properties of Wine Pigments and Tannins. *Am J Enol Vitic* **2006**, 57, 298–305, doi:10.5344/ajev.2006.57.3.298.

4. Downey, M.O.; Dokoozlian, N.K.; Krstic, M.P. Cultural Practice and Environmental Impacts on the Flavonoid Composition of Grapes and Wine: A Review of Recent Research. *Am J Enol Vitic* **2006**, 57, 257–268, doi:10.5344/ajev.2006.57.3.257.

5. Esteban, M.A.; Villanueva, M.J.; Lissarrague, J.R. Effect of Irrigation on Changes in the Anthocyanin Composition of the Skin of cv Tempranillo (*Vitis vinifera* L) Grape Berries during Ripening. *J Sci Food Agric* **2001**, 81, 409–420, doi:10.1002/1097-0010(200103)81:4<409::AID-JSFA830>3.0.CO;2-H.

6. Lee, J.; Skinkis, P.A. Oregon 'Pinot Noir' Grape Anthocyanin Enhancement by Early Leaf Removal. *Food Chem* **2013**, 139, 893–901, doi:10.1016/j.foodchem.2013.02.022.

7. Ryan, J.M.; Revilla, E. Anthocyanin Composition of Cabernet Sauvignon and Tempranillo Grapes at Different Stages of Ripening. *J Agric Food Chem* **2003**, 51, 3372–3378, doi:10.1021/jf020849u.

8. Pascual, O.; Ortiz, J.; Roel, M.; Kontoudakis, N.; Gil, M.; Gómez-Alonso, S.; García-Romero, E.; Canals, J.M.; Hermosín-Gutíerrez, I.; Zamora, F. Influence of Grape Maturity and Prefermentative Cluster Treatment of the Grenache Cultivar on Wine Composition and Quality. *OENO One* **2016**, 50, 169–181, doi:10.20870/oeno-one.2016.50.4.1824.

9. Federico Casassa, L.; Beaver, C.W.; Mireles, M.S.; Harbertson, J.F. Effect of Extended Maceration and Ethanol Concentration on the Extraction and Evolution of Phenolics, Colour Components and Sensory Attributes of Merlot Wines. *Aust J Grape Wine Res* **2013**, 19, 25–39, doi:10.1111/ajgw.12009.

10. Pérez-Navarro, J.; García Romero, E.; Gómez-Alonso, S.; Izquierdo Cañas, P.M. Comparison between the Phenolic Composition of Petit Verdot Wines Elaborated at Different Maceration/Fermentation Temperatures. *Int J Food Prop* **2018**, 21, 996–1007, doi:10.1080/10942912.2018.1479856.

11. Pontallier, P. The Intervention of Oak Wood in the Making of Great Red Wines. *J Wine Res* **2007**, 3, 241–247, doi:10.1080/09571269208717935.

12. Morata, A.; Gómez-Cordovés, M.C.; Suberviola, J.; Bartolomé, B.; Colomo, B.; Suárez, J.A. Adsorption of Anthocyanins by Yeast Cell Walls during the Fermentation of Red Wines. *J Agric Food Chem* **2003**, 51, 4084–4088, doi:10.1021/jf021134u.

13. Kontoudakis, N.; Esteruelas, M.; Fort, F.; Canals, J.M.; De Freitas, V.; Zamora, F. Influence of the Heterogeneity of Grape Phenolic Maturity on Wine Composition and Quality. *Food Chem* **2011**, 124, 767–774, doi:10.1016/j.foodchem.2010.06.093.

14. Ignat, I.; Volf, I.; Popa, V.I. A Critical Review of Methods for Characterisation of Polyphenolic Compounds in Fruits and Vegetables. *Food Chem* **2011**, 126, 1821–1835, doi:10.1016/j.foodchem.2010.12.026.

15. dos Santos, C.A.T.; Páscoa, R.N.M.J.; Porto, P.A.L.S.; Cerdeira, A.L.; González-Sáiz, J.M.; Pizarro, C.; Lopes, J.A. Raman Spectroscopy for Wine Analyses: A Comparison with Near and Mid Infrared Spectroscopy. *Talanta* **2018**, 186, 306–314, doi:10.1016/j.talanta.2018.04.075.

16. Lambrecht, K.; Nieuwoudt, H.; du Toit, W.; Aleixandre-Tudo, J.L. Moving towards In-Line Monitoring of Phenolic Extraction during Red Wine Fermentations Using Infra-Red Spectroscopy Technology. Influence of Sample Preparation and Instrumentation. *J Food Compos Anal* **2022**, 110, 104542, doi:10.1016/j.jfca.2022.104542.

17. Lambrecht, K.; Nieuwoudt, H.; du Toit, W.; Aleixandre-Tudo, J.L. Optimisation of PLS Calibrations for Filtered and Untreated Samples towards In-Line Monitoring of Phenolic Extraction during Red-Wine Fermentations. *Fermentation* **2022**, 8, 231, doi:10.3390/fermentation8050231/S1.

18. Iland, P. Techniques for Chemical Analysis and Quality Monitoring During Winemaking; Patrick Iland, Ed. Wine Promotions, **2000**; ISBN 064638435X.

19. Mercurio, M.D.; Dambergs, R.G.; Herderich, M.J.; Smith, P.A. High Throughput Analysis of Red Wine and Grape Phenolics - Adaptation and Validation of Methyl Cellulose Precipitable Tannin Assay and Modified Somers Color Assay to a Rapid 96 Well Plate Format. *J Agric Food Chem* **2007**, 55, 4651–4657, doi:10.1021/JF063674N.

20. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

21. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA Simultaneous Component Analysis: A Tutorial Review. *Anal Chim Acta X* **2020**, 6, 100061, doi:10.1016/j.acax.2020.100061.

22. Nadal, M. Phenolic Maturity in Red Grapes. In *Methodologies and Results in Grapevine Research*; Delrot, S., Medrano, H., Or, E., Bavaresco, L., Grando, S., Eds.; Springer Netherlands, **2010**; pp. 389–409 ISBN 9789048192830.

23. Canals, R.; Llaudy, M.C.; Valls, J.; Canals, J.M.; Zamora, F. Influence of Ethanol Concentration on the Extraction of Color and Phenolic Compounds from the Skin and Seeds of Tempranillo Grapes at Different Stages of Ripening. *J Agric Food Chem* **2005**, 53, 4019–4025, doi:10.1021/jf047872v.

24. Kelebek, H.; Canbas, A.; Selli, S.; Saucier, C.; Jourdes, M.; Glories, Y. Influence of Different Maceration Times on the Anthocyanin Composition of Wines Made from *Vitis vinifera* L. cvs. Boğazkere and Öküzgözü. *J Food Eng* **2006**, 77, 1012–1017, doi:10.1016/j.jfoodeng.2005.08.032.

25. Sims, C.A.; Bates, R.P. Effects of Skin Fermentation Time on the Phenols, Anthocyanins, Ellagic Acid Sediment, and Sensory Characteristics of a Red *Vitis rotundifolia* Wine. *Am J Enol Vitic* **1994**, 45, 56–62, doi:10.5344/ajev.1994.45.1.56.

26. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

27. Suriano, S.; Alba, V.; Tarricone, L.; Di Gennaro, D. Maceration with Stems Contact Fermentation: Effect on Proanthocyanidins Compounds and Color in Primitivo Red Wines. *Food Chem* **2015**, 177, 382–389, doi:10.1016/j.foodchem.2015.01.063.

28. Fanzone, M.; Peña-Neira, A.; Gil, M.; Jofré, V.; Assof, M.; Zamora, F. Impact of Phenolic and Polysaccharidic Composition on Commercial Value of Argentinean Malbec and Cabernet Sauvignon Wines. *Food Res Int* **2012**, 45, 402–414, doi:10.1016/j.foodres.2011.11.010.

29. Schorn-García, D.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Methodologies Based on ASCA to Elucidate the Influence of a Subprocess: Vinification as a Case of Study. *J Chemom* **2022**, e3465, doi:10.1002/CEM.3465.

30. González-Caballero, V.; Sánchez, M.T.; Fernández-Novales, J.; López, M.I.; Pérez-Marín, D. On-Vine Monitoring of Grape Ripening Using Near-Infrared Spectroscopy. *Food Anal Methods* **2012**, 5, 1377–1385, doi:10.1007/S12161-012-9389-3.

31. Schorn-García, D.; Giussani, B.; García-Casas, M.J.; Rico, D.; Martin-Diana, A.B.; Aceña, L.; Busto, O.; Boqué, R.; Mestres, M. Assessment of Variability Sources in Grape Ripening Parameters by Using FTIR and Multivariate Modelling. *Foods* **2023**, 12, 962, doi:10.3390/foods12050962.

32. Moura, J.C.M.S.; Bonine, C.A.V.; de Oliveira Fernandes Viana, J.; Dornelas, M.C.; Mazzafera, P. Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants. *J Integr Plant Biol* **2010**, 52, 360–376, doi:10.1111/j.1744-7909.2010.00892.x.

33. Fourie, E.; Aleixandre-Tudo, J.L.; Mihnea, M.; du Toit, W. Partial Least Squares Calibrations and Batch Statistical Process Control to Monitor Phenolic Extraction in Red Wine Fermentations under Different Maceration Conditions. *Food Control* **2020**, 115, 107303, doi:10.1016/j.foodcont.2020.107303.

References (Section 2)

1. Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Dubourdieu, D. Handbook of Enology, The Chemistry of Wine: Stabilization and Treatments: Second Edition; Wiley, **2006**; Vol. 2; ISBN 9780470010396.

2. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

3. Schorn-García, D.; Giussani, B.; García-Casas, M.J.; Rico, D.; Martin-Diana, A.B.; Aceña, L.; Busto, O.; Boqué, R.; Mestres, M. Assessment of Variability Sources in Grape Ripening Parameters by Using FTIR and Multivariate Modelling. *Foods* **2023**, 12, 962, doi:10.3390/foods12050962/.

4. Jackson, D.I.; Lombard, P.B. Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. *Am J Enol Vitic* **1993**, 44, 409–430, doi:10.5344/ajev.1993.44.4.409.

5. Tempere, S.; Pérès, S.; Espinoza, A.F.; Darriet, P.; Giraud-Héraud, E.; Pons, A. Consumer Preferences for Different Red Wine Styles and Repeated Exposure Effects. *Food Qual Prefer* **2019**, 73, 110–116, doi:10.1016/j.foodqual.2018.12.009

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Section 3

Spatially offset Raman spectroscopic (SORS) analysis of wine alcoholic fermentation.
A preliminary study

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Process Analytical Technologies (PAT)

Alcoholic fermentation, although primarily known for the conversion of glucose and fructose into ethanol, is a highly complex process that includes several biochemical transformations and reactions. This is because the consumption of metabolites by yeast and other microorganisms to carry out their metabolism, as well as the production of "quorum sensing" metabolites as part of the microorganism's response to the presence of other microorganisms, contribute even more to the complexity of this process. Beyond ethanol production, the fermentation process plays a crucial role in the development of the organoleptic characteristics of wine, including the release of primary aroma compounds present in their glycosylated form or the production of secondary aroma compounds as a result of yeast metabolism. Moreover, the extraction yield of colour and polyphenolic compounds from grape seeds and skins is also related to this process as it is facilitated by the increasing amount of ethanol[1].

To achieve a specific and quality product, the oenologist must keep all these processes under control, since a failure could lead to deviations in both the physicochemical and organoleptic properties of the wine produced[2]. When such deviations happen, they often involve complex changes, so off-site laboratory analyses are typically required to evaluate the situation, since the necessary analytical techniques are usually not available in cellars. However, this leads to a delay in getting results, subsequently affecting the decision-making process[3]. Even minor compounds that due to the deviation may be generated in low concentrations can significantly influence the organoleptic properties of the wine. Therefore, any waiting time that can be minimized in the analysis process can contribute to preserving the desired organoleptic properties and overall quality of the wine. The more time that passes there is an increased likelihood of generation of undesirable compounds, which could have low detection thresholds and have a noticeable impact on the final product.

Recognizing the importance of closely monitoring and controlling the fermentation process to ensure the desired final product quality, industries have shifted towards implementing Process Analytical Technologies (PAT) and adopting Quality by Design (QbD) principles[4]. PAT, as defined by the American Food and Drug Administration (FDA), refers to a system that incorporates timely measurements of critical quality and performance attributes of raw and in-process materials and processes during manufacturing[5]. It has been widely

employed in the pharmaceutical industry to enhance product quality and process efficiency[4,6], but its principles perfectly align with the main goals of the wine industry. By employing PAT, winemakers can gain real-time insights into the fermentation process, allowing for proactive adjustments and interventions to maintain product quality throughout the entire production cycle, rather than relying on limited information obtained at the final stages. Considering all these benefits and potential for improvement, there is no doubt that the wine industry should adopt these methodologies to achieve consistent quality and optimize production processes.

Different types of measurement can be made when applying PAT approaches, so it is possible to distinguish:

- At-line measurements: in this mode the equipment, which may be a benchtop or a portable or handheld device, is near to the process. Minimal or no sample pre-treatment is needed to successfully apply it in a fast way[7].

- On-line measurements: in this mode the sample is diverted or measured through an aperture covered by a non-absorptive material, assembled in the production line[8].

- In-line measurements: This mode involves placing the sensor directly in the process, typically submerged, to perform real-time analysis under actual process conditions[9].

The choice of the measurement approach depends on the specific needs and objectives of the PAT implementation. At-line measurements are commonly used in the initial stages to serve as a bridge between off-line and in-process methodologies, providing real-time information without the need for modifying process conditions or equipment[10].

Regardless of the approach used, the implementation of PAT requires a comprehensive understanding of the properties and components of raw materials, their evolution throughout the process, the quality parameters of the final product and the potential factors that may influence every step. By employing the PAT approach, many data will be acquired so statistical treatments will be necessary to stablish suitable and reliable relationships between the obtained signals or data and the desired properties at different

stages of the process[11]. This holistic approach ensures a thorough assessment of the entire production chain.

Among the different techniques proposed as part of PAT methodologies, the ones that provide fingerprint signals in a simply and fast way are the most useful. Thus, spectroscopy techniques, such as NIR, MIR and Raman, have a high demand due to their ability to provide comprehensive information about the process and its properties making a measurement in seconds and in a respectful way with the environment. By using NIR, MIR and Raman spectroscopy, a wide range of compounds can be analysed, enabling the monitoring of processes and obtaining valuable insights[12]. However, Raman spectroscopy offers several advantages over other spectroscopic techniques, particularly due to the weak Raman scattering exhibited by water and ethanol molecules. This property proves highly valuable in the analysis of alcoholic fermentation, where water and ethanol are the main components[13].

Raman spectroscopy

Raman spectroscopy is a light scattering technique that uses a monochromatic high intensity light that interacts with molecular bonds within a material, causing them to scatter this incident light. The occurrence of Raman scattering is rare, with only about 1 in every 10^8 photons undergoing this process. Raman scattering occurs when the initial energy state of a molecule differs from its final energy state as the molecule relaxes. The final energy state can be higher (Stokes shift) or lower (anti-Stokes shift) than the initial state[14]. The different energy excitation and relaxation processes are represented in Figure 1.

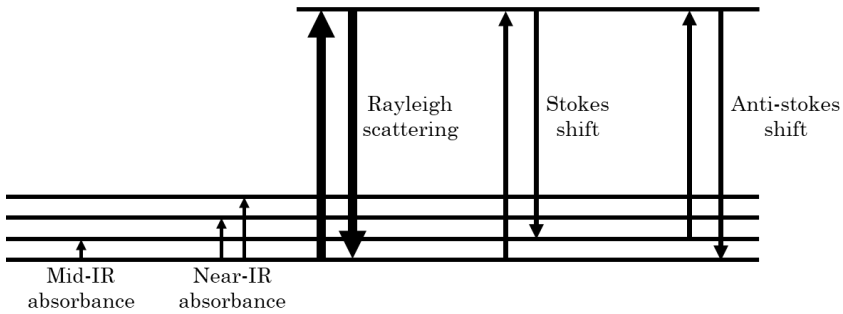


Figure 1. Energy-level diagram of mid-IR, near-IR, Rayleigh, and Raman scattering. Adapted from Rodriguez-Saona *et al.*[13].

The significant advances in Raman spectrometers in recent years have expanded the applicability of this technique in PAT methodologies. The development of portable Raman spectrometers and the introduction of techniques like Spatially Offset Raman Spectroscopy (SORS) have further enhanced its capabilities. SORS combines traditional Raman spectroscopy with spatially offset optics to enable chemical analysis of samples even when located beneath a material or surface. The key difference with conventional Raman spectroscopy lies in the collection of scattered signals. In conventional Raman, the incident light and detector are parallel and located in the same position, obtaining the Raman spectra from the point of incidence. In SORS, the detector is shifted relative to the light sources, allowing photons from deeper layers within the sample to be captured. As a result, two signals are obtained: one from the surface and another from the inner layers or subsurface. Although the latter signal exhibits a lower intensity, it is suitable for chemical analysis. In the case of SORS, it becomes essential to "enhance" the collected signal, photon by photon, using mechanical and digital signal amplification, to acquire the desired spectroscopic information[15,16].

SORS has demonstrated its efficacy in various applications, particularly in obtaining chemical information from samples packed or wrapped in different materials such as skin, paper, plastic, and glass[16–19]. This technique is particularly valuable in mitigating fluorescence interference by disregarding potential fluorescence emanating from the container material. Additionally, the use of a high-wavelength laser further aids in reducing the impact of fluorescence on the analysis, enabling accurate and reliable spectral measurements[15]. The combination of Raman spectroscopy and SORS within the realm of PAT provides researchers and industries with powerful tools for non-destructive analysis, process monitoring and quality control in the food and beverage industry.

On-line Spatially Offset Raman Spectroscopy in alcoholic fermentation

This is the first time the application of SORS technology has been investigated for monitoring alcoholic fermentation. SORS opens up new possibilities for wineries to incorporate QbD and PAT approaches. However, to gain a comprehensive understanding of the variations caused by different sources in the application of SORS, it was necessary to compare the results obtained when working with at-line and on-line measurement setups, as well as to study the effect of the presence of yeast in the fermentation media. In

order to assess these differences objectively, various measurement configurations and geometries were tested.

According to the approach previously described in the **Tutorial**, plotting Principal Component 1 (PC1) or PC2 against time allowed for visualizing the temporal evolution of the spectra. The same strategy was applied to Raman spectra across different measurement configurations[20]: at-line (taking out a sample and analysing it with and without centrifugation), and on-line (measuring through the fermentation container from two different angles: horizontal plane and vertical plane). To quantify the effect of each factor studied, ASCA was employed. The use of ASCA enabled the assessment of the effects associated with different measurement setups and the presence of yeast, providing valuable insights into the variations observed in the SORS measurements.

Figure 2. Experimental set-up used for Raman spectra collection.

It should be noted that there might be external factors that influence the excitation of molecules, as there is a limited occurrence of Raman scattering. The impact of light sources such as sunlight, light bulbs, or fluorescent tubes becomes crucial as they can potentially affect the reproducibility of the spectra. To address this issue in our study, particular attention was given to the collection of Raman spectra by considering various light sources. Prior to analysis, a preventive measure was implemented by switching off the light sources for a duration of 5 minutes. This step was taken because artificial light sources can continue to emit photons even after being switched off, which could interfere with the accuracy of the spectra. The experimental setup employed for Raman spectroscopy is depicted in Figure 2, where a box was used with a suitable sized aperture to accommodate the sampling device of the SORS spectrometer. This controlled setup ensured standardized

conditions for data acquisition and minimized the influence of external light sources on the Raman measurements.

The results from the feasibility of using SORS to monitor alcoholic fermentation have been compiled in **Paper 4**. The on-line approach allowed for successful monitoring through the fermentation container, eliminating the need to take a sample. Additionally, the presence of yeast suspended in the media did not affect the results. All these results demonstrate the potential of SORS technology to be used in both at-line and on-line setups, even in wine-making conditions with suspended yeast. Building upon these findings, a new avenue of research can be pursued, focusing on the effective application of SORS not only for monitoring but also for controlling the process of alcoholic fermentation. This opens up prospects for enhancing the quality and efficiency of the fermentation process in winemaking.

Spatially offset Raman spectroscopic (SORS) analysis of wine alcoholic fermentation. A preliminary study

Daniel Schorn–García[a], Jokin Ezenarro[a], Laura Aceña[a], Olga Busto[a], Ricard Boqué[b], Barbara Giussani[c], Montserrat Mestres[a]

[a]Universitat Rovira i Virgili, Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, 43007 Tarragona, Spain

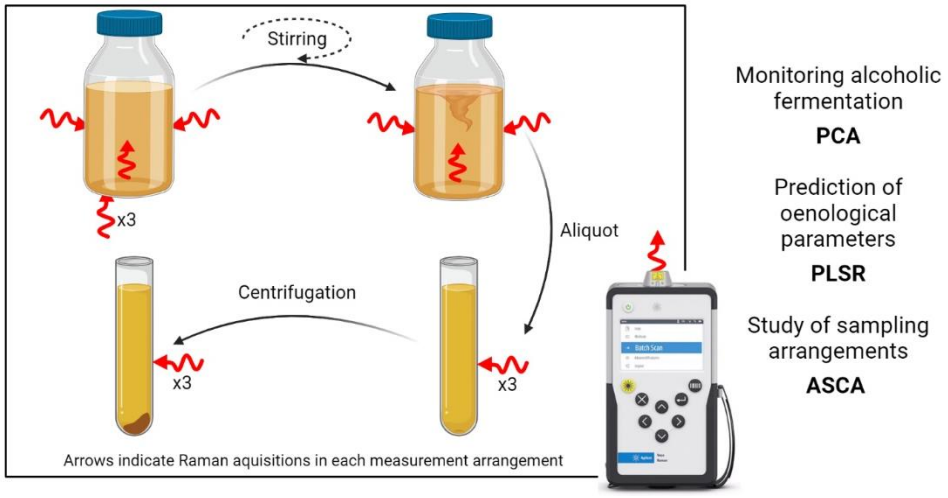
[b]Universitat Rovira i Virgili. Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, 43007 Tarragona, Spain

[c]Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, 22100 Como, Italy

Fermentation 9(2), 115 (2023)

Abstract

Spatially offset Raman spectroscopy (SORS) is a non-invasive analytical technique that allows the analysis of samples through a container. This makes it an effective tool for studying food and beverage products, as it can measure the sample without being affected by the packaging or the container. In this study, a portable SORS equipment was used for the first time to analyse the alcoholic fermentation process of white wine. Different sample measurement arrangements were tested in order to determine the most effective method for monitoring the fermentation process and predicting key oenological parameters. The best results were obtained when the sample was directly measured through the glass container in which the fermentation was occurring. This allowed the accurate monitoring of the process and the prediction of density and pH with a root mean square error of cross-validation (RMSECV) of 0.0029 $g \cdot L^{-1}$ and 0.04, respectively, and R^2 values of 0.993 and 0.961 for density and pH, respectively. Additionally, the sources of variability depending on the measurement arrangements were studied using ANOVA-Simultaneous Component Analysis (ASCA). This study shows that the SORS technique can be a valuable tool to help winemakers monitor fermentation in real time, allowing them to make decisions to readjust the process if necessary. Furthermore, the obtained results provide valuable information for future applications of the studied technique.



Keywords

Process Analytical Technologies (PAT); multivariate analysis; infrared spectroscopy; analysis through packaging

Introduction

Winemaking is a dynamic biochemical process, in which microorganisms and chemical compounds affect the course of the process and the properties and quality of the final product. The main reaction of the winemaking process is the transformation of sugars, basically glucose and fructose, into ethanol and carbon dioxide. This reaction is the basis of the alcoholic fermentation and occurs as part of yeast metabolism[1]. Moreover, due to secondary metabolism pathways other yeast metabolites are released at different concentrations, which are related to the organoleptic and physicochemical properties of the final product[1]. This process is highly sensitive to variations in chemical composition and external conditions (e.g. pH, sugar concentration or temperature), which could even lead to stuck or sluggish fermentations[2]. For this reason, close monitoring of the process is necessary to obtain real-time information that allows the necessary corrective measures to be taken in time to avoid a quality loss of the final product[3].

Currently, most wineries do not carry out exhaustive fermentation controls, but only rely on daily measurements of density and pH, as well as an organoleptic assessment by the oenologist[4]. However, the information provided by these parameters is not sufficient when abnormal fermentation behaviour appears and the decisions to be made depend on the results of more complex analyses, usually performed at-line or in external laboratories. This implies a delay in obtaining information and, therefore, in the application of corrective measures[5]. This delay can be a big problem if these actions are applied once the unwanted chemical substances have already impacted the organoleptic properties of the final product[6]. In this context, Process Analytical Technologies (PAT) are increasingly used, as they allow product quality to be assured through real-time measurements throughout the process[7]. Among the technologies, vibrational spectroscopy is frequently used in the food and beverage industries as it brings together several positive aspects: it allows obtaining information of each molecule in the matrix; it is sustainable, eco-friendly and requires minimal or no sample processing[5].

Near-infrared (NIR) and Mid-infrared (MIR) spectroscopies have already been used to monitor fermentation[8,9] and even to successfully detect deviations of the process[10,11]. However, these technologies have drawbacks in their application, such as the complexity of the signal in the case of NIR, as a consequence of weak overtones and combination bands; and the high MIR absorption of water bonds that make the use of special sampling

devices mandatory to overcome signal saturation[12]. Raman spectroscopy could overcome these drawbacks as water bonds do not produce signal saturation and this spectroscopy produces sharp peaks that can be associated to specific bonds[12,13]. Even these advantages it has rarely been used in on-line wine monitoring[14] mainly due to the fact that Raman spectroscopy is reliant on transparency of containers for the analysis of the sample contained within.

Nowadays, with Spatially Offset Raman Spectroscopy (SORS) it is possible to acquire Raman signals through many millimetres of different materials, allowing the analysis of the sample through barriers such as plastic, glass, paper, etc. The main characteristic of this type of measurement is that, unlike conventional Raman, the irradiation of the laser (excitation) and the collection of light do not coincide geometrically, but rather a spatially displaced measurement is performed. Therefore, since excitation and detection are spatially separate, a balanced subtraction of the two measurements creates a clean spectrum of the material contained[15,16]. This makes SORS a non-invasive, non-destructive technique that does not require sample preparation[17] and, thanks to the availability of commercial portable equipments, it becomes of great interest as a tool for process quality control.

The aim of this research has been to evaluate the use of a portable spatially offset Raman spectrophotometer (SORS) as a monitoring tool for wine alcoholic fermentation. Different sample measurement configurations were tested to assess the performance of the Raman spectrophotometer. Principal Component Analysis (PCA), Partial Least Squares Regression (PLSR) and ANOVA-Simultaneous Component Analysis (ASCA) were used to monitor the process, predict the main oenological parameters and study the sources of variability in Raman spectra.

Materials and Methods

Fermentation Samples

Four alcoholic fermentations were carried out at small-scale (microfermentations) into 2.0 L glass cylindrical containers. White must from the dilution of a commercial concentrated must was used. The concentrated white must (Julián Soler S.A., Cuenca, Spain) was stored at −20 ºC to avoid any biochemical or chemical evolution. 24 hours before its use it was kept at 4 ºC to defrost it. Then it was diluted with MilliQ water to adjust the sugar

concentration to 200 $g \cdot L^{-1}$. To ensure a sufficient yeast assimilable nitrogen, the diluted must was supplemented with 0.30 $g \cdot L^{-1}$ of Actimabio* (Agrovin S.A., Ciudad Real, Spain) and ENOVIT® (Spindall S.A.R.L., Gretz Armainvilliers, France), respectively.

For each one of the four microvinifications (or biological replicates) 1.5 mL of diluted must was inoculated with commercial *Saccharomyces cerevisiae* yeast (Viniferm Revelación, Agrovin S.A.) ensuring an initial yeast population of $3 \cdot 10^6$ $CFU \cdot mL^{-1}$. Before inoculation, rehydration of the yeast was done following the suppliers' instructions. All microvinifications were kept at a constant temperature of 20 ºC until the end of alcoholic fermentation (that is, until the density was lower than 0.995 $g \cdot L^{-1}$, which is equivalent to a final sugar content under 1 $g \cdot L^{-1}$).

The alcoholic fermentation process was routinely controlled by measuring directly into the container density and pH, twice a day, with a portable densimeter (Densito2Go, Mettler Toledo, United States) and a portable pH meter with a 201 T electrode (7+ series portable pH-meter, XS Instruments, Italy). Both apparatus were calibrated once a day before use with a reference standard of 0.9982 $g \cdot mL^{-1}$ and two reference standards of pH 7.00 and 4.00, respectively. The densimeter cell and pH-meter electrode were thoroughly cleaned with deionized water before each new reading.

Raman Analysis

The Vaya Raman (Agilent, California, USA) portable SORS equipment was used in this research. This device uses a laser with an excitation wavelength of 830 nm to seek the suppression of fluorescence. The power of the laser was automatically adjusted by the spectrometer (taking into account the signal detected) to 450 mW reaching the maximum power available. The spectra were acquired in a range from 350 to 2000 cm^{-1}. The equipment performs two consecutive measurements: with zero offset (parallel pathway of the laser and the detector) and with spatial offset (shift of 0.7 mm from the point of the laser incidence to the detector). After internal processing, the equipment records the SORS spectrum, which is the result of a scaled subtraction of the two measurements that provides a clean spectrum of the sample without the influence of the container layers.

The sample was measured from two different angles to study the performance of the SORS spectrophotometer in each case: directly next to the fermentation glass container in the horizontal plane (three equidistant points) and at the bottom of the container in the

vertical plane. Furthermore, to detect the possible effect of turbidity, the samples were also measured directly next to the container in the horizontal plane after vigorous stirring. In every case, the SORS technology was applied to avoid the effect of the container. A scheme of the different analysis arrangements is shown in Figure 1.

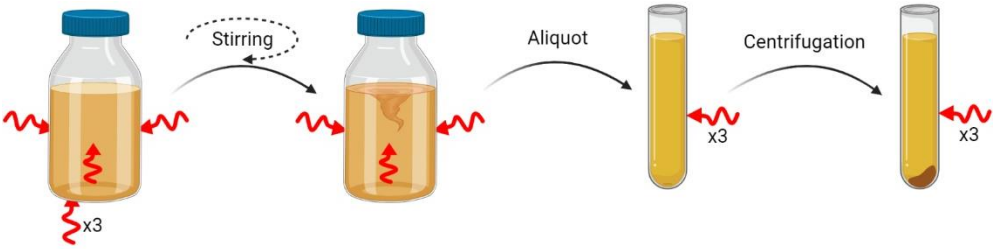


Figure 1. Scheme of analysis arrangement. Each red arrow indicates the position where Raman spectrometer was placed.

Aliquots of samples were also collected during the fermentation and analysed before and after centrifugation to obtain their spectra (without the influence of the container). Raman spectra of aliquots were obtained using a vial-mode configuration (Figure 1). The samples, contained in glass vials, were inserted into the spectrometer sample holder and the light was focused on the sample to perform the scans. Measurement times ranged from 30 seconds to 2 minutes per sample, while laser exposure times ranged from 0.5 to 2.0 seconds.

Each of the four microfermentations or biological replicates was analysed in triplicate in each measurement arrangement (shown in Figure 1) twice a day (with a twelve-hour gap) for seven consecutive days, which accounts for 14 sampling points for each biological replicate in each of the measurement arrangement.

Data Analysis

Spectral Data Pre-Processing

Spectra were imported to MATLAB (R2021a, 9.10; MathWorks, Natick, MA) to create five datasets (one for each analysis configuration) of 168 spectra (triplicates of 56 samples (4 biological replicates x 14 sampling points)), each consisting of 1651 wavenumbers. Multivariate analysis were performed using the PLS Toolbox (v9.0; Eigenvector Research

Inc., Earglerock, USA). Different pre-processing combinations were tested to overcome noise observed in the raw spectra: wavelet denoising (as implemented in Wavelet Toolbox v6.0), Savitzky-Golay (SG) smoothing and Standard Normal Variate (SNV), as implemented in the PLS Toolbox. The best spectral pre-processing combination to build all the multivariate models was wavelet denoising. Data were mean-centred before modelling.

Multivariate Data Analysis

Principal Component Analysis (PCA) is a well-known exploratory technique used to visualize the data (see e.g. esbensen *et al.*[18] and references therein). It allows to detect groups and trends among samples and identify samples outliers, and to study relationships between the investigated variables. The algorithm decomposes the data into a set of new variables called Principal Components, which are linear combinations of the original variables, orthogonal between them and retaining most of the information, and an error matrix, that contains non-relevant information and model noise. The projections of the samples onto the new space of Principal Components, known as scores, and the angles between the original variables and the principal components, known as loadings, can be plotted to reveal trends between samples and to help understand the sources of variability in the data. Partial Least Squares Regression (PLSR) was used to build models to predict density and pH[19]. For each measurement arrangement, an X data matrix (containing the sample spectra) and a Y data matrix (containing two columns, one for density and the other for pH), were used. Model evaluation was carried out using Cross-Validation as the validation method (ten random subsets iterated ten times) and calculating the Root Mean Squared Error of Cross-Validation (RMSECV) statistic (Equation 1). This parameter describes how well the model predicts new samples using a cross validation strategy[12]. First, a partition of the set is done, and each subset is excluded from the model building and then the excluded subset is predicted. This process is repeated until every subset is used.

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_i^{n_t}(y_{t,i} - \hat{y}_{t,i})^2}{n_t}} \qquad \text{Equation 1}$$

$\hat{y}_{t,i}$ are the densities or pHs predicted by the models, $y_{t,i}$ are the measured values (actual pH or density), and n_t is the number of samples in the cross-validation set. RMSECV is an

approximation of the average error to be expected in future predictions when the calibration model is applied to unknown samples.

Two statistical dimensionless parameters were used to assess model's predictive ability: Ratio of Performance to Deviation (RPD) and Range Error Ratio (RER) (Equation 2 and 3)[20]:

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSECV}} \qquad \text{Equation 2}$$

$$\mathrm{RER} = \frac{y_{max} - y_{min}}{\mathrm{RMSECV}} \qquad \text{Equation 3}$$

SD is the standard deviation of the reference parameters (pH and density), and y_{max} and y_{min} are the maximum and minimum vales of the reference parameters. Model predictions were considered good enough when RPD and RER are greater than two and ten, respectively.

Finally, Analysis of Variance (ANOVA)-Simultaneous Component Analysis (ASCA) was used to decompose the variability sources affecting the data. ASCA is a multivariate extension of ANOVA, which decomposes the variation in the data into the main effects and their binary combinations, obtained from a predefined experimental design[21]. In this study, three variability factors were considered: suspended yeast (stirring or no stirring in direct measurements; or centrifuge or no centrifuge in aliquot measurements), the alcoholic fermentation process itself and the biological replicates, and the interactions between them. The first step of ASCA requires partitioning the centred X matrix according to equation 4:

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = \mathbf{X}_{\text{Ripening}} + \mathbf{X}_{\text{pos.plant}} + \mathbf{X}_{\text{pos.bunch}} + \mathbf{X}_{\text{Ripening x pos.plant}} + \mathbf{X}_{\text{Ripening x pos.bunch}} + \mathbf{X}_{\text{pos.plant x pos.bunch}} + \mathbf{X}_{\text{res}} \qquad \text{(Equation 4)}$$

where 1 is a vector of ones, $\mathbf{m}^T$ is the average spectrum of the samples, $\mathbf{X}_{\text{suspended yeast}}$, $\mathbf{X}_{\text{alcoholic fermentation}}$, $\mathbf{X}_{\text{biological replicates}}$ are the matrices of the main factors, $\mathbf{X}_{\text{suspended yeast x alcoholic fermentation}}$, $\mathbf{X}_{\text{suspendended yeast x biological replicate}}$, $\mathbf{X}_{\text{alcoholic fermentation x biological replicate}}$ are the effect matrices for the binary interactions, and $\mathbf{X}_{\text{res}}$ is the residual matrix that collects the variability not taken into account in the experimental design. Each matrix is centred and contains the mean profiles of the samples corresponding to each factor or interaction level.

Then, the matrices are decomposed using a technique called Simultaneous Component Analysis (SCA). It is important to note that SCA can be thought of as a form of principal component analysis that is constrained by ANOVA[21].

Results and Discussion

Evolution of the alcoholic fermentation

The correct progress of alcoholic fermentation was checked by two daily measurements of density and pH. The evolution of the density (Figure 2a) showed the typical sigmoidal trend[9], reaching the end of the fermentation in around 156 hours. pH also exhibited the typical trend during alcoholic fermentation (Figure 2b) with a decrease until the midpoint of the tumultuous fermentation (at 76 hours) due to the consumption of nitrogenous compounds and the release of organic acids. The second part of the alcoholic fermentation showed a slight increase in pH[1]. As all biological replicates behaved similarly, they were all considered as fermentations under control, and thus their Raman spectra were used for further analysis.

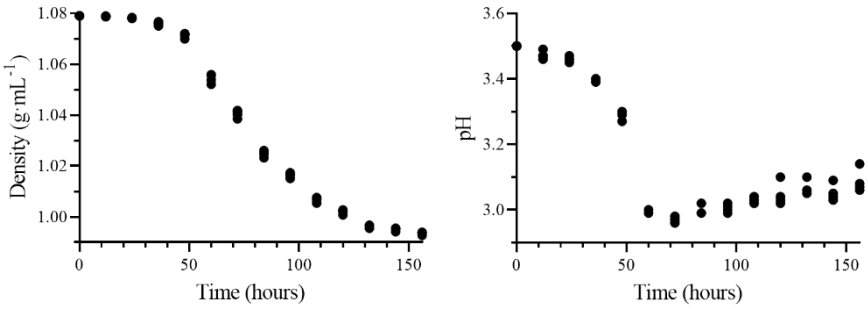


Figure 2. Evolution of (a) density and (b) pH during alcoholic fermentation.

Raman monitoring of the alcoholic fermentation

The first objective of the study was to establish the optimal arrangement for the acquisition of the sample signal, taking into consideration the angle between the Raman sensor and the sample container and the turbidity influence of the suspended yeast in the alcoholic fermentation matrix. For horizontal measurements through the fermentation glass container, three replicates were measured at three different side points of the container (always at the same three equidistant points) before and after manual stirring.

Additionally, before stirring, a vertical measurement was performed at the bottom of the container, also in triplicate and always at the same point, to evaluate the effect of the yeast deposited at the bottom on the signal obtained.

To obtain a spectrum of the sample without the possible container interference, direct analysis of an aliquot of the fermentation samples was performed both before and after centrifugation.

To evaluate the influence of the container in the SORS spectra of the samples, an aliquot of each sampling point was analysed in a glass tube container provided by the manufacturer and that meets the provisions required for Raman spectroscopy. For this analysis, and to evaluate the effect of the suspended yeast on the signal as it was done for the direct analysis, measurements were performed before and after centrifugation. Three replicates of the sample of each microvinification at each sampling time were measured.

Figure 3 shows an example of the SORS spectra obtained at the beginning and at the end of the fermentation process. The main peaks in the first stage of alcoholic fermentation could be assigned to the vibration of different bonds belonging to the sugar molecules: 451 cm^{-1} for δ(C–C–O), 521 cm^{-1} for cyclic carbons, 1124 cm^{-1} for angular torsion and CH_2 group at 1455 cm^{-1}. Other important bands described in the alcoholic fermentation are C–O–H bending at 1424 cm^{-1}, C–C stretching at 1130 cm^{-1} and C–O stretching at 1072 cm^{-1} [22]. As for the spectrum corresponding to the end of fermentation, the most characteristic band of ethanol assigned to the C–C stretching vibration around 880 cm^{-1} can be clearly seen[12,23].

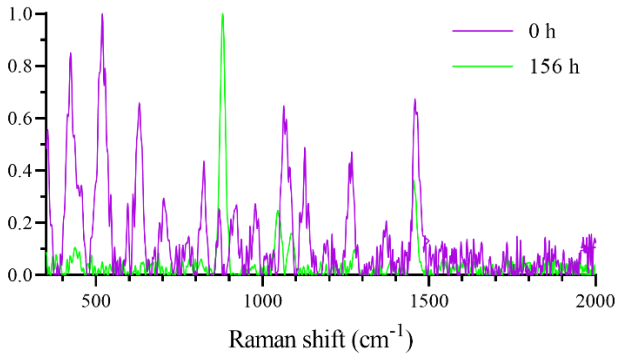


Figure 3. SORS spectra of one biological replicate of alcoholic fermentation at the beginning (purple) and at the end (green) of the process.

To detect possible different trends depending on the arrangement used for signal acquisition, a PCA model was built for each configuration using the mean spectra of the three replicates. The dimensions of the matrices for each analysis configuration were 4 biological replicates x 14 time points x 1651 wavenumbers, which were unfolded to obtain two-dimensional matrices of 56 x 1651.

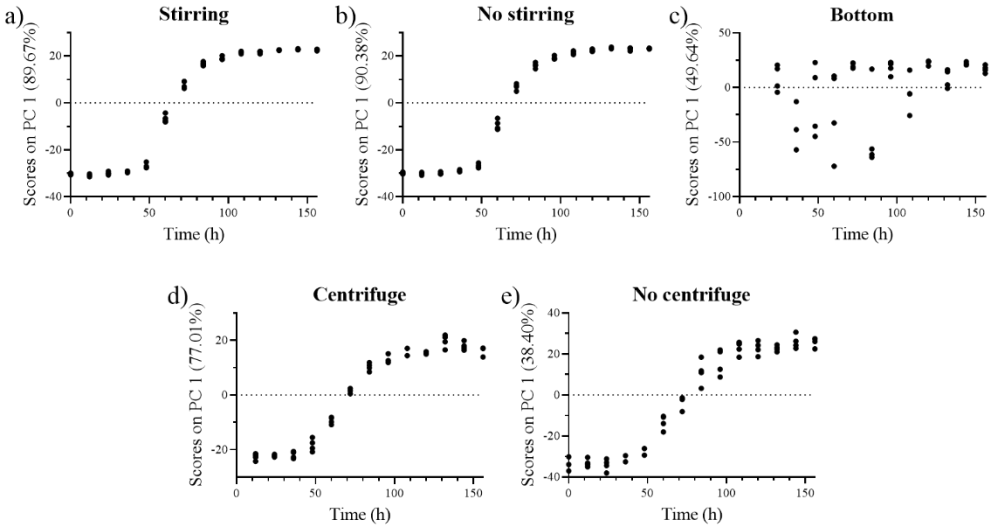


Figure 4. Score plots of the first Principal Component (PC1) of each PCA model for the different measurement configurations. a-c) PC1 versus time obtained from SORS spectra of direct measures in the fermentation glass container. d-e) PC1 versus time obtained from Raman spectra of aliquots of the fermentation.

As shown in Figure 4, the evolution of the scores in the first principal component (PC1) over time showed a process-related sigmoidal curve for each signal acquisition arrangement, except for the vertical measurement at the bottom of the sample container. The sigmoidal behaviour observed in relation to density evolution allowed the monitoring of the process as established in the laboratory using mid-infrared spectroscopy and outlined in a previously published tutorial[9]. Regarding the process, after 96 hours, a stabilization of the score values was observed, meaning that the first PC was not able to distinguish final stages of alcoholic fermentation, that end at hours 156 as previously mentioned (see paragraph 3.1). However, the second PC for "stirring" and "no stirring" PCA models showed an evolution over time in that period (Figure S1). This could be related to the final part of the consumption of sugars and the production of ethanol, as the loadings of this component were also related to sugars and ethanol. When looking at the reference samples (centrifuged and not centrifuged aliquots) they showed small differences in the

score values of the first PC of the replicates at the end of the process, which could be related to biological variability. However, both configurations showed the same sigmoidal shape, and could be used to monitor the fermentation. Finally, in all cases the variance explained by the first PC ranged from 38.40 to 90.38% depending on the type of signal acquirement, showing a great variability compared to other variability sources in the spectra.

Prediction of oenological parameters

As both density and score values of the first PC showed a sigmoidal evolution, PLSR was used to predict density values during the alcoholic fermentation process. The five data matrices above described (56 samples x 1651 variables) were used in the experiment. The model performance parameters are shown in Table 1. RPD and RER were used to compare the performance of the prediction models. In addition, the number of LVs to be considered was optimized based on the curve of the RMSECV.

Table 1. Density prediction results in each analysis configuration. LV: Latent Variables used in the PLSR models, RMSECV: Root Mean Square Error of Cross-Validations (expressed in $g \cdot mL^{-1}$), R^2: determination coefficient, RPD: Ratio of Performance to Deviation, RER: Range Error Ratio.

	LV	**RMSECV** ($g \cdot mL^{-1}$)	**R^2**	**RPD**	**RER**
Side analysis without stirring	2	0.0029	0.993	11.8	29.8
Side analysis after stirring	2	0.0030	0.992	11.5	28.8
Bottom analysis	1	0.0285	0.204	1.2	3.0
Aliquot analysis without centrifuge	3	0.0041	0.986	8.4	21.0
Aliquot analysis after centrifuge	4	0.0042	0.985	8.2	20.5

As can be seen from the results, similar prediction models were obtained for “stirring” and “no stirring”, and between aliquot with and without centrifuge. The analysis at the bottom of the container, as expected by the inspection of the PCA scores in different PCs, was not able to predict density. This was also confirmed by looking at the loadings of the first LV of every measurement arrangement, with a large peak at 880 cm^{-1}, being the major peak associated to ethanol (Figure S2a). Comparable results have been published in the literature for the prediction of sugars, as density is mainly related to the concentration of sugars in the fermenting must[12,14,22,23]. Both RPD and RER indicated that the best models

were the ones of the direct analysis in the side of the container both before and after stirring. This was consistent with the literature stating that SORS coupled to chemometrics overcomes some Raman problems such as fluorescence[16]. In addition, regarding the present study these better results were attributed to the greater representativeness of the fermentation process when analysing the whole fermentation sample than when taking only an aliquot.

Due to its importance in alcoholic fermentation monitoring, PLSR was also applied to predict pH. Following the same methodology as for density, the five data matrices (56 x 1651) were used to predict pH in every measurement arrangement. The quality parameters of the different models are summarized in Table 2.

Table 2. pH prediction results in each analysis configuration. LV: Latent Variables used in the PLSR models, RMSECV: Root Mean Square Error of Cross-Validations, R^2: determination coefficient, RPD: Ratio of Performance to Deviation, RER: Range Error Ratio.

	LV	**RMSECV**	**R^2**	**RPD**	**RER**
Side analysis without stirring	2	0.04	0.961	5.0	13.5
Side analysis after stirring	2	0.04	0.953	5.0	13.5
Bottom analysis	3	0.12	0.483	1.7	4.5
Aliquot analysis without centrifuge	3	0.07	0.888	2.8	7.7
Aliquot analysis after centrifuge	3	0.08	0.806	2.5	6.8

As in the case of density, the performances of the pH prediction models were different for each signal acquisition configuration. The best performances were obtained with measurements performed at the container side, poorer performances were obtained when analysing the aliquots and the poorest prediction ability was obtained when measuring at the bottom of the container. For every predictive model, the loadings of the LVs (Figure S2b-c) were related to sugars, ethanol, and also a peak at 1760 cm^{-1} related to C=O stretching[16]. Regarding the quality parameters of the best models, the results were in agreement with the literature when using on-line conventional (non SORS) Raman spectroscopy[12,23], and proved the usefulness of SORS for monitoring the alcoholic fermentation.

Figure 5 shows the plots of predicted vs measured values for the best models, using the spectra obtained from the measurements made directly on the side of the container without previous stirring.

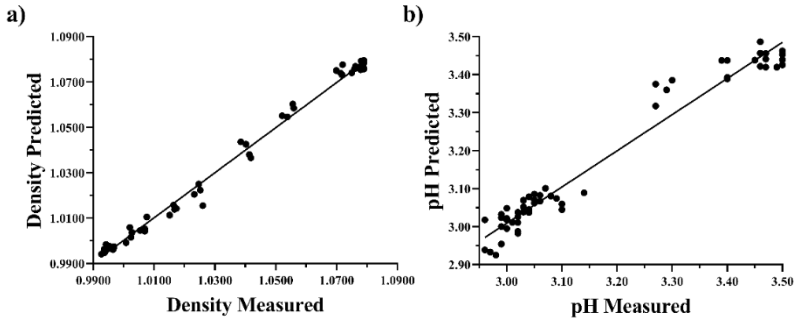


Figure 5. Predicted *vs* measured values for the best prediction models of (a) density (expressed in $g \cdot mL^{-1}$) and (b) pH. Both models were calculated using spectra obtained from measurements without stirring.

Variability sources

Two ASCA models were calculated to study the variability in the spectra associated to the presence of suspended yeast. Two extended matrices of 112 (56 samples x 2 measurement arrangements) x 1651 wavenumbers were used. ASCA results are expressed in terms of % Effect, which indicates the contribution of each factor to the matrix variability. A permutation test of 10000 iterations was performed in each ASCA model to assess the significance of each factor (a *p*-value under 0.05 means the factor is significant)[24]. The ASCA results for the measurements made directly at the container side before and after stirring are summarized in Table 3.

As can be seen, the most relevant factor was the alcoholic fermentation process, which has already been shown to be very important in ATR-FTIR data of wine alcoholic fermentation[25]. This was consistent with the fact that during alcoholic fermentation yeast consumes approximately 200 $g \cdot L^{-1}$ of sugars and produces 13% ethanol, being the main source of variability in the data. As far as sample agitation was concerned, the results showed that stirring the container to resuspend the yeast in the medium prior to the spectrum acquisition had no significant influence, making this step unnecessary. The biological replicate did not have a significant effect, but the interaction between biological replication and alcoholic fermentation process did. This could be explained by small differences in the speed of the process, with an effect of 2.08%.

Table 3. ASCA results for the "stirring" and "no stirring" measurement arrangements, showing the percentage of variance (Effect (%)) for each factor and the *p*-value obtained from the permutation test. A *p*-value < 0.05 means the factor is significant.

Factor	Effect (%)	*p*-value
Stirring	0.05	0.3597
Alcoholic Fermentation process	95.39	0.0001
Biological replicate	0.15	0.3255
Stirring x Alcoholic Fermentation	0.56	0.4431
Stirring x biological replicate	0.11	0.7845
Alcoholic fermentation x biological replicate	2.08	0.0020

The ASCA results for the measurements of the aliquot before and after centrifugation are summarized in Table 4.

Table 4. ASCA results for centrifuge and no centrifuge measurement disposition, showing the percentage of variance (Effect (%)) for each factor and then *p*-value resulting of the permutation test. A *p*-value < 0.05 means the factor is significant.

Factor	Effect (%)	*p*-value
Centrifugation	0.42	0.1477
Alcoholic Fermentation	71.08	0.0001
Biological replicate	0.45	0.6289
Centrifugation x Alcoholic Fermentation	6.69	0.0001
Centrifugation x biological replicate	0.30	0.4271
Alcoholic fermentation x biological replicate	18.18	0.0001

ASCA results for aliquot measurements confirmed that the alcoholic fermentation factor is again the greatest source of variability found in the data. However, in the case of the aliquot, the course of the alcoholic fermentation was different for the biological replicates, as the interaction between both factors was found large and significant. Alcoholic fermentation batches without centrifugation showed more variability in each sampling point than batches with centrifugation. The standard deviation of the scores for each sampling point in Section 3.2 PCA ranged from 0.44 to 4.74 and 1.45 to 11.55 for aliquots with and without centrifugation, respectively. This result also explained why the interaction between centrifugation and alcoholic fermentation factors was also found large

and significant. This ASCA results for aliquot analysis were in concordance with the results in Section 3.2 and 3.3, showing that centrifugation had a significant impact in the acquisition of Raman spectra, as in a short pathway with a priori no dispersion effects yeast may have caused light scattering.

Conclusions

The results presented demonstrate the ability of SORS to obtain the Raman spectra of wine alcoholic fermentation in a rapid, non-invasive and non-destructive way. Different measurement strategies were tested: through the container where the fermentation was occurring with and without previous stirring and taking an aliquot to be analysed in the Raman sample holder. The variability associated to the different measurement configurations was assessed using ASCA, showing that manual stirring has no impact in the performance of the SORS. Contrary, when an aliquot is taken from the container, the centrifugation has an impact due to the removal of suspended yeast. However, all measurement configurations, except those taken at the bottom of the container, successfully monitored the fermentation process and predicted density and pH as the main oenological parameters used in winemaking. The results obtained allow establishing that SORS equipment could be used in the control of fermentation processes in an on-line configuration. Further research should be conducted to study the prediction of other oenological parameters of interest and the detection of process deviations, as well as the potential use of SORS for the analysis of other types of wine and alcoholic beverages.

Funding

Grant PID2019-104269RR-C33 funded by MCIN/AEI/ 10.13039/501100011033. This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-García, D.). Grant URV Martí i Franqués - Banco Santander (2021PMF-BS-12).

Acknowledgments

The authors would like to thank Agilent Technologies Spain for providing the portable Raman spectrophotometer used in this study. Graphical abstract was done using BioRender.com

References (Paper 4)

1. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

2. Urtubia, A.; Hernández, G.; Roger, J.M. 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.

3. Emparán, M.; Simpson, R.; Almonacid, S.; Teixeira, A.; Urtubia, A. Early Recognition of Problematic Wine Fermentations through Multivariate Data Analyses. *Food Control* **2012**, 27, 248–253, doi:10.1016/j.foodcont.2012.03.029.

4. Bisson, L.F. Stuck and Sluggish Fermentations. *Am J Enol Vitic* **1999**, 50, 107–119.

5. Cozzolino, D. Advantages, Opportunities, and Challenges of Vibrational Spectroscopy as Tool to Monitor Sustainable Food Systems. *Food Anal Methods* **2022**, 15, 1390–1396, doi:10.1007/s12161-021-02207-w.

6. 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, 282–297, doi:10.1080/05704928.2015.1132721.

7. U.S. Department of Health and Human Services Food and Drug Administration Guidance for Industry Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance; **2004**.

8. 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, 947–955, doi:10.1007/s00217-010-1227-5.

9. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.

10. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Monitoring Wine Fermentation Deviations Using an ATR-MIR Spectrometer and MSPC Charts. *Chemometr Intell Lab* **2020**, 201, 104011, doi:10.1016/j.chemolab.2020.104011.

11. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

12. Teixeira dos Santos, C.A.; Páscoa, R.N.M.J.; Porto, P.A.L.S.; Cerdeira, A.L.; González-Sáiz, J.M.; Pizarro, C.; Lopes, J.A. Raman Spectroscopy for Wine Analyses: A Comparison with near and Mid Infrared Spectroscopy. *Talanta* **2018**, 186, 306–314, doi:10.1016/j.talanta.2018.04.075.

13. Cozzolino, D. Sample Presentation, Sources of Error and Future Perspectives on the Application of Vibrational Spectroscopy in the Wine Industry. *J Sci Food Agric* **2015**, 95, 861–868, doi:10.1002/jsfa.6733.

14. Ávila, T.C.; Poppi, R.J.; Lunardi, I.; Tizei, P.A.G.; Pereira, G.A.G. Raman Spectroscopy and Chemometrics for On-Line Control of Glucose Fermentation by *Saccharomyces cerevisiae*. Biotechnol Prog **2012**, 28, 1598–1604, doi:10.1002/btpr.1615.

15. Pérez-Beltran, C.H.; Pérez-Caballero, G.; Andrade, J.M.; Cuadros-Rodríguez, L.; Jiménez-Carvelo, A.M. Non-Targeted Spatially Offset Raman Spectroscopy-Based Vanguard Analytical Method to Authenticate Spirits: White Tequilas as a Case Study. *Microchem J* **2022**, 183, 108126, doi:10.1016/j.microc.2022.108126.

16. Jiménez-Carvelo, A.M.; Arroyo-Cerezo, A.; Bikrani, S.; Jia, W.; Koidis, A.; Cuadros-Rodríguez, L. Rapid and Non-Destructive Spatially Offset Raman Spectroscopic Analysis of Packaged Margarines and Fat-Spread Products. *Microchem J* **2022**, 178, 107378, doi:10.1016/j.microc.2022.107378.

17. Arroyo-Cerezo, A.; Jimenez-Carvelo, A.M.; González-Casado, A.; Koidis, A.; Cuadros-Rodríguez, L. Deep (Offset) Non-Invasive Raman Spectroscopy for the Evaluation of Food and Beverages-A Review. *LWT-Food Sci Technol* **2021**, 149, 111822, doi:10.1016/j.lwt.2021.111822.

18. Esbensen, K.H.; Geladi, P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. *Comprehensive Chemometrics* **2009**, 2, 211–226, doi:10.1016/b978-044452701-1.00043-0.

19. Geladi, P.; Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal Chim Acta* **1986**, 185, 1–17, doi:10.1016/0003-2670(86)80028-9.

20. Fearn, T. Assessing Calibrations: SEP, RPD, RER and R^2. *NIR news* **2002**, 13, 12–13, doi:10.1255/nirn.689.

21. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

22. Wu, Z.; Xu, E.; Long, J.; Wang, F.; Xu, X.; Jin, Z.; Jiao, A. Measurement of Fermentation Parameters of Chinese Rice Wine Using Raman Spectroscopy Combined with Linear and Non-Linear Regression Methods. *Food Control* **2015**, 56, 95–102, doi:10.1016/j.foodcont.2015.03.015.

23. Fuller, H.; Beaver, C.; Harbertson, J.; Jordão, M.; Michelle Wine Estates, S. Alcoholic Fermentation Monitoring and PH Prediction in Red and White Wine by Combining Spontaneous Raman Spectroscopy and Machine Learning Algorithms. *Beverages* **2021**, 7, 78, doi:10.3390/beverages7040078.

24. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA Simultaneous Component Analysis: A Tutorial Review. *Anal Chim Acta X* **2020**, 6, 100061, doi:10.1016/j.acax.2020.100061.

25. Schorn-García, D.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Methodologies Based on ASCA to Elucidate the Influence of a Subprocess: Vinification as a Case of Study. *J Chemom* **2022**, e3465, doi:10.1002/CEM.3465.

Supplementary Material

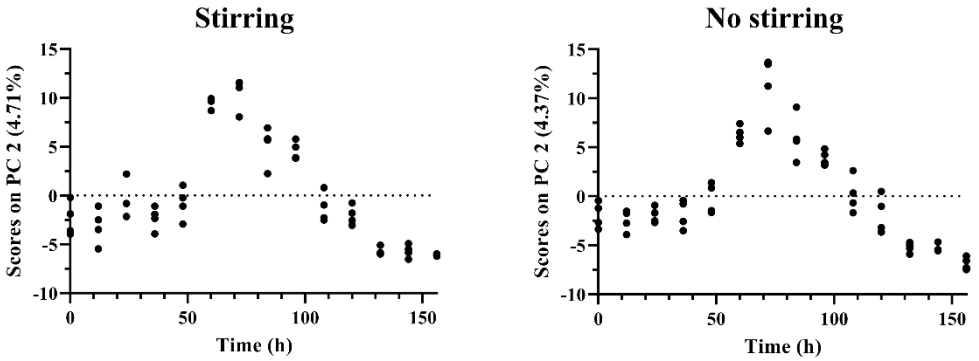


Figure S1. Scores plots of the second Principal Component (PC2) of stirring and no stirring PCA model.

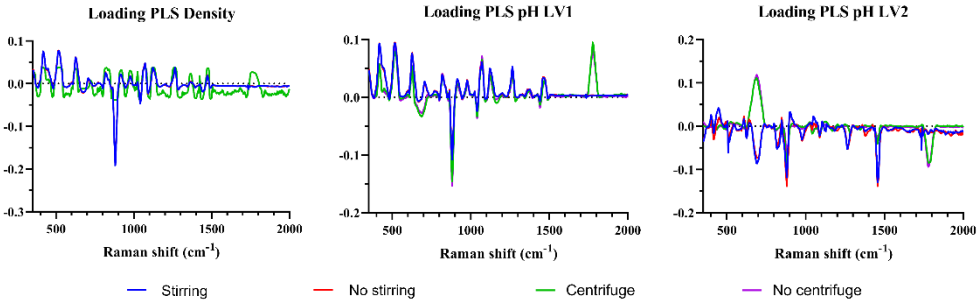


Figure S2. Loading plot of density (first LV) and pH (first and second) PLSR model. Different colours mean different analytical methodologies (blue – Stirring, red – No stirring, green – Centrifuge, purple – No centrifuge).

References

1. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

2. Menezes, J.C. Process Analytical Technology in Bioprocess Development and Manufacturing. In *Comprehensive Biotechnology*, Second Edition; Moo-Young, M., Ed.; Academic Press, **2011**; Vol. 3, pp. 501–509 ISBN 9780080885049.

3. Mello, L.D.; Kubota, L.T. Review of the Use of Biosensors as Analytical Tools in the Food and Drink Industries. *Food Chem* **2002**, 77, 237–256, doi:10.1016/S0308-8146(02)00104-8.

4. Martins Fukuda, I.; Francini, C.; Pinto, F.; Dos, C.; Moreira, S.; Morais Saviano, A.; Lourenço, F.R. Design of Experiments (DoE) Applied to Pharmaceutical and Analytical Quality by Design (QbD). *J Pharm Sci* **2018**, 54, doi:10.1590/s2175-97902018000001006.

5. Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacuring, and Quality Assurance; Services, U.S. Department of Health and Human, Food and Drug Administration, **2004**.

6. Benito, M.T.J.; Ojeda, C.B.; Rojas, F.S. Process Analytical Chemistry: Applications of Near Infrared Spectrometry in Environmental and Food Analysis: An Overview. *Appl Spectrosc Rev* **2008**, 43, 452–484, doi:10.1080/05704920802031382.

7. Roychoudhury, P.; Harvey, L.M.; 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, 218–224, doi:10.1016/j.aca.2006.01.037.

8. Wang, Q.; Li, Z.; Ma, Z.; Liang, L. Real Time Monitoring of Multiple Components in Wine Fermentation Using an On-Line Auto-Calibration Raman Spectroscopy. *Sens Actuators B Chem* **2014**, 202, 426–432, doi:10.1016/j.snb.2014.05.109.

9. Claßen, J.; Aupert, F.; Reardon, K.F.; Solle, D.; Scheper, T. Spectroscopic Sensors for In-Line Bioprocess Monitoring in Research and Pharmaceutical Industrial Application. *Anal Bioanal Chem* **2016** 409(3), 651–666, doi:10.1007/s00216-016-0068-x.

10. Kueppers, S.; Haider, M. Process Analytical Chemistry-Future Trends in Industry. *Anal Bioanal Chem* **2003**, 376, 313–315, doi:10.1007/S00216-003-1907-0.

11. van den Berg, F.; Lyndgaard, C.B.; Sørensen, K.M.; Engelsen, S.B. Process Analytical Technology in the Food Industry. *Trends Food Sci Technol* **2013**, 31, 27–35, doi:10.1016/j.tifs.2012.04.007.

12. 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. In *Measurement, Monitoring, Modelling and Control of Bioprocesses. Advances in Biochemical Engineering/Biotechnology*; Mandenius, C., Titchener-Hooker, N., Eds.; Springer, Berlin, Heidelberg, **2013**; Vol. 132, pp. 249–280 ISBN 9783642368370.

13. Rodriguez-Saona, L.; Ayvaz, H.; Wehling, R.L. Infrared and Raman Spectroscopy. In *Food Analysis*; Nielsen, S.S., Ed.; Springer, Cham, **2017**; pp. 107–127, doi: 10.1007/978-3-319-45776-5_8.

14. Boyaci, I.H.; Temiz, H.T.; Geniş, H.E.; Acar Soykut, E.; Yazgan, N.N.; Güven, B.; Uysal, R.S.; Bozkurt, A.G.; İlaslan, K.; Torun, O. Dispersive and FT-Raman Spectroscopic Methods in Food Analysis. *RSC Adv* **2015**, 5, 55606–55624, doi:10.1039/c4ra12463d.

15. Parker, A.W. Spatially Offset Raman Spectroscopy. In *Encyclopedia of Spectroscopy and Spectrometry*; Lindon, J.C., Tranter, G.E., Koppenaal, D.W., Eds.; Academic Press, **2017**; pp. 143–148, doi: 10.1016/b978-0-12-409547-2.12151-1.

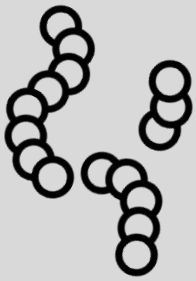
16. Arroyo-Cerezo, A.; Jimenez-Carvelo, A.M.; González-Casado, A.; Koidis, A.; Cuadros-Rodríguez, L. Deep (Offset) Non-Invasive Raman Spectroscopy for the Evaluation of Food and Beverages-A Review. *LWT-Food Sci Tech* **2021**, 149, 111822, doi:10.1016/j.lwt.2021.111822.

17. Jiménez-Carvelo, A.M.; Arroyo-Cerezo, A.; Bikrani, S.; Jia, W.; Koidis, A.; Cuadros-Rodríguez, L. Rapid and Non-Destructive Spatially Offset Raman Spectroscopic Analysis of Packaged Margarines and Fat-Spread Products. *Microchem J* **2022**, 178, 107378, doi:10.1016/j.microc.2022.107378.

18. Pérez-Beltran, C.H.; Pérez-Caballero, G.; Andrade, J.M.; Cuadros-Rodríguez, L.; Jiménez-Carvelo, A.M. Non-Targeted Spatially Offset Raman Spectroscopy-Based Vanguard Analytical Method to Authenticate Spirits: White Tequilas as a Case Study. *Microchem J* **2022**, 183, 108126, doi:10.1016/j.microc.2022.108126.

19. Arroyo-Cerezo, A.; Jiménez-Carvelo, A.M.; González-Casado, A.; Ruisánchez, I.; Cuadros-Rodríguez, L. The Potential of the Spatially Offset Raman Spectroscopy (SORS) for Implementing Rapid and Non-Invasive in-Situ Authentication Methods of Plastic-Packaged Commodity Foods-Application to Sliced Cheeses. *Food Control* **2023**, 146, 109522, doi:10.1016/j.foodcont.2022.109522.

20. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.



Section 4

Methodologies based on ASCA to elucidate the influence of a subprocess: vinification as a case of study

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Wine spoilage

The waxy layer on the grape skin, known as the grape bloom, is a natural habitat for a diversity of microorganisms, including yeasts and bacteria. Among the bacterial populations that can be found, the microorganism groups with oenological relevancy relevance are acetic acid bacteria (AAB) and lactic acid bacteria (LAB). The variety and abundance of bacteria are influenced by various factors such as climatic conditions and the sanitary state of the grapes. In particular, AAB populations tend to thrive in more humid climates and in situations where grapes are damaged or broken, providing them with metabolic access to grape sugars. During the harvest, these bacteria can be transferred to wineries through the grape bunches and other sources[1].

AAB are classified as aerobic bacteria, meaning that they require oxygen-rich environments for their growth and survival[2]. Oxygen acts as an electron acceptor in their metabolic processes. However, during alcoholic fermentations, where the primary microorganism involved is yeast, the matrix conditions shift to anaerobic environments due to the production of carbon dioxide. This lack of oxygen induces the AAB population, if initially low, to remain relatively stable or even decrease in number[3].

Many factors can influence the risk of AAB population raising, such as an excessive aeration of must or wine or the use of unsanitary grapes with broken skins. A high AAB population is considered as a spoilage in winemaking due to its primary metabolic reaction, which involves the oxidation of ethanol to acetic acid. This reaction gives wine a characteristic pungent vinegar-like smell. Acetic acid contributes significantly to volatile acidity and for this reason, in the wine industry, volatile acidity is a regulated parameter used to assess quality, with a requirement for it to be maintained below a specified legal threshold (less than 1.2 $g \cdot L^{-1}$ or lower depending on Protected Designation of Origin legal specifications). It is worth noting that in certain cases, a minor presence of acetic acid can actually enhance the overall aroma profile of the wine, adding complexity and character. However, the sensory threshold to notice the pungency of acetic acid is considerably less than the legal limit, at 0.5 or 0.8 $g \cdot L^{-1}$ for white and red wines, respectively. In addition to acetic acid production, AAB metabolism also produces other compounds such as acetaldehyde, ethyl acetate, and diacetyl, which can further impact the sensory attributes of the wine[1,2,4]. Depending on the desired characteristics of the final wine, this would be appropriate or not for the oenologist.

Considering the influence of AAB on wine quality, it is crucial to employ effective analytical methods capable of accurately determining the concentration of acetic acid produced by AAB, particularly at early stages. This proactive approach allows winemakers to take necessary measures to prevent excessive acetic acid accumulation, ensuring the wine meets quality standards and regulatory requirements. Spectroscopy proves to be highly suitable in this context, as it provides a rapid and reliable means of obtaining comprehensive information about the chemical composition of wine samples. Moreover, the portability of spectroscopic instruments enables their integration into wine production processes.

Data matrix treatment

The application of spectroscopy to follow the fermentation process produces a vast amount of data. Multiple variables are measured, either wavelengths or wavenumbers, every two, four, eight, or more depending on the spectral resolution. To better characterize the fermentation process, numerous batches are considered, and as they are sampled over time, a third dimension of data is obtained. This three-way matrix takes the form of Batches x Variables x Time, with dimensions *I*x*J*x*K*. However, in order to effectively apply usual chemometric techniques such as PCA or PLS, a two-way matrix is required[5]. A schematic representation of the different approaches to reduce the dimensionality of data and used in the chapter are shown in Figure 1.

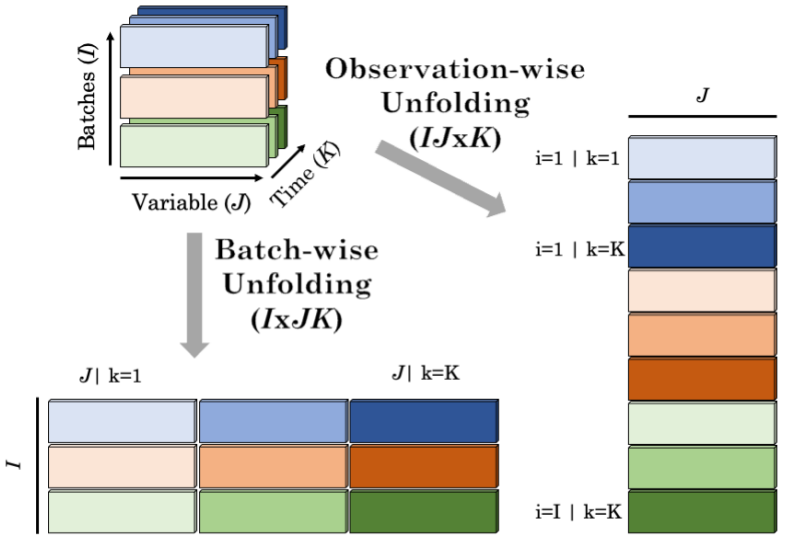


Figure 1. Schematic representation of a possible unfolding or reorganization of a three-way matrix.

The organization or unfolding of this three-dimensional matrix has a significant impact on the resulting model, as the data structure can highlight different variances and covariances within the dataset. Various unfolding approaches can be employed, including Batch-wise unfolding and Observation-wise unfolding.

Batch-wise unfolding involves unfolding the variable and time dimensions while maintaining the batch direction. In this approach, each row contains the data from a specific batch across all measurements. This unfolding technique enables the extraction of relationships between variables of different batches or even within the same batch in different moments of the process. By including all the data from a batch within a row, both the within-time and over-time correlated information of the variables is captured. It is important to note that the same batch length is required for modelling batches, especially for future batches[6].

In turn, Observation-wise unfolding maintains the variable direction while unfolding the batch and time dimensions. In this approach, each row corresponds to the measurements of a specific batch at a particular time point. This unfolding technique allows for the inclusion of batches with missing values and accommodates differences between batches in the time length of the process[7].

The choice of the unfolding approach depends on the specific goals and characteristics of the process being analysed, for instance if the process has to be evaluated as a whole (batch-wise approach) or if each measurement has to be evaluated individually to visualize its evolution (observation-wise approach). Both approaches offer unique insights into the correlations and patterns of the data, providing valuable information for modelling and understanding the fermentation process.

Spectral pre–processing and variable selection

Data pre-processing is a crucial step to transform raw data into a more suitable form for further analysis and interpretation. Pre-processing involves a series of techniques that aim to enhance the quality of data and remove unwanted noise and artifacts[8]. These artifacts can arise from various sources, such as fluctuations in instrument settings, environmental factors, or inherent limitations of the measurement technique. By eliminating them, the pre-processing stage ensures that the subsequent data analysis

focuses on the true underlying patterns and relationships within the data, improving the reliability of the models[9,10].

Normalization and scaling are among the techniques used in pre-processing to adjust the variability of measured variables. Normalization (such as SNV or MSC) brings the data to a common scale, which is particularly useful when variables have different units or ranges. Scaling (such as standardization), on the other hand, aims to achieve a comparable magnitude for variables, preventing any single variable from dominating the analysis due to its larger values. Other pre-processing techniques such as smoothing or derivatives (such as Savitzky-Golay smoothing and derivative) focus on the reduction of noise in the spectra and the enhancement of small peaks that may contain useful information of low concentration compounds[11].

The choice of an optimal pre-processing technique or a combination of techniques is a critical decision that depends on the specific properties of the data and the objectives of the analysis. Factors such as the nature of the data (such as Raman or MIR, which behaves differently), the presence of outliers or noise, the distribution of the variables, and the desired analytical outcomes all influence the selection of appropriate pre-processing techniques. Careful consideration and evaluation are necessary to ensure that the chosen techniques effectively enhance the data quality without introducing any unintended bias or distortion[12].

In addition to pre-processing, variable selection plays a crucial role in optimizing the accuracy and performance of spectroscopic methods. While pre-processing techniques aim to enhance the overall quality of the data, variable selection focuses on identifying and retaining the most relevant variables that contribute significantly to the model. Even with correct pre-processing, certain regions of the spectra may not contain critical information about the underlying process. These regions have been assigned with less weight in the model. However, if unrelated noise is present in these regions, it can have a detrimental effect on the model performance. Therefore, removing variables in which noise overrides relevant information often leads to better performance of spectroscopic methods[13].

Variable selection can be performed using different approaches, such as forward or backward selection. In forward selection, variables are added to the model one by one or as a set until a satisfactory model performance is achieved. Conversely, in backward selection, variables are progressively subtracted from the dataset until a desirable level of

model performance error is achieved. These selection methods help refine the model by identifying the most informative variables and discarding those that contribute less to the analysis[14,15].

Moreover, the analyst's prior knowledge can be valuable in manually selecting variables to remove, based on their chemical properties. Spectroscopy, particularly in the mid-infrared region, exhibits specific regions associated with particular chemical bonds. Taking advantage of this knowledge, analysts can focus on the selection of regions of interest that are known to contain relevant information about the process under investigation[16].

By combining pre-processing techniques with variable selection techniques, researchers can refine the data, reduce noise, and choose the most informative variables for analysis. Using all the available tools, a model for the training samples is obtained. This approach not only enhances the accuracy and performance of spectroscopic methods but also allows for a deeper understanding of the chemical properties and regions of interest in the spectra[12].

Acetic acid bacteria spoilage

As mentioned above, achieving a good model requires multiple steps in data processing and analysis. This is especially crucial when studying complex biochemical processes like alcoholic fermentation, with underlying subprocesses such as acetification or AAB spoilage. These processes take distinct roles in the evolution of the overall process over time, with alcoholic fermentation being dominant in magnitude compared to acetification. However, even taking place in a less extent, the acetification process significantly influences the organoleptic properties of wine.

In the third study of this thesis different approaches were used to effectively study the weight and impact of acetification. Thus, to evaluate the different approaches ASCA was used, which decomposes the data matrix into its constituent factors, enabling the quantification of both processes. Using ASCA, it became possible to objectively evaluate the influence of both alcoholic fermentation and acetification on the spectral data. Therefore, pre-processing techniques, variable selection procedures, and unfolding strategies were tested to enhance the spectroscopic signal of the unwanted process, the AAB metabolism.

The findings of our research on the possible strategies for studying alcoholic fermentation and acetification processes are presented in **Paper 5**. As explained in it, the selection of specific segments of the process and targeted parts of the spectra generated the most promising results. In particular, these selected segments aimed to mitigate the influence of alcoholic fermentation on the spectra, mainly the spectral bands associated with sugars and ethanol. A division of the process in several stages was also aimed to reduce the influence of the phase of the process characterized by the highest sugars to ethanol conversion rate. Through the application of these results, it is possible to propose a methodology to improve the spectroscopic quality for process control strategies of a subprocess partially hidden by a majoritarian process.

Methodologies based on ASCA to elucidate the influence of a subprocess: vinification as a case of study

Daniel Schorn–García[a], Barbara Giussani[b], 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, 43007 Tarragona, Spain

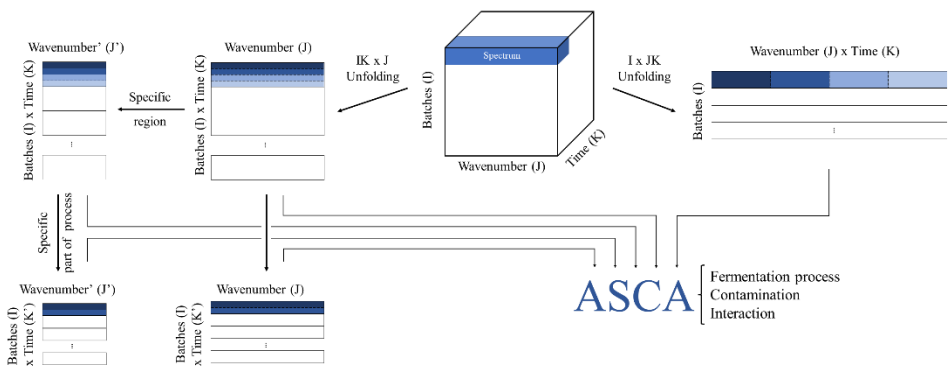
[b]Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, 22100 Como, Italy

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

Journal of Chemometrics, e3465 (2023)

Abstract

In food manufacturing and processing, food matrix complexity usually makes it difficult to detect unwanted subprocesses, which can impact the quality of the final product. In the case of wine alcoholic fermentation, the main process is the conversion of sugars into ethanol and carbon dioxide, but the presence of some unwanted microorganisms could lead to wine contamination by production of undesired minor compounds. In the study we present, an intentional contamination of the vinification process by the addition of acetic acid bacteria was studied using a portable Fourier Transform Infrared (FT–IR) spectrometer. ANOVA Simultaneous Component Analysis (ASCA) was used to unravel these minor variability sources. However, as the subprocess is two orders of magnitude lower in concentration than the main process, different methodologies were used to enhance the ASCA results, such as to select a specific spectral region related to acetic acid bacteria metabolism, to divide the process in time intervals related to the different phases, or to unfold the data matrix in different ways. In addition, spectral pre-processing was optimized to scale up small peaks related to the subprocess. Our results show that several methodologies to build ASCA models can be applied to emphasize and better characterize bacteria contamination subprocesses.



Keywords

portable FTIR, acetic acid bacteria, ANOVA Simultaneous Component Analysis, process deviation

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 products and processes[1]. The reason is that these techniques are very fast and allow the simultaneous determination of several parameters with minimal or no sample pre-treatment. Thus, once developed, implemented and properly combined with multivariate data analysis (MVDA), they allow obtaining a lot of information from a sample almost immediately[2].

In the wine sector, correct monitoring of alcoholic fermentation is essential to ensure the production of high-quality wines, since inefficient management of this process would lead to the production of undesired molecules with a negative impact on the quality, both organoleptic and chemical, of the final product[3]. To achieve good control of alcoholic fermentation, it is necessary to perform daily measurements of different physicochemical parameters, which ensure the correct development of the process[4]. However, most of these measurements are time-consuming and require specific laboratory equipment and trained personnel, which forces many wineries to 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 the process[4], would be highly beneficial for those wineries that cannot have their own analytical laboratory to perform at-line analyses. Even for wineries with their own laboratory, PAT strategies are helpful, as the frequency of analysis would increase substantially, and pre-treatment of the samples would be considerably simplified[5].

A key step in food monitoring is data collection, which is usually time-consuming[6]. In addition, when choosing techniques such as spectroscopy to monitor process parameters, reference techniques must be also implemented during the model construction phase, making the whole procedure in some cases expensive as well.

MVDA techniques allows to extract useful information from spectroscopic data and to correlate it with the reference values through a proper multivariate model[7]. The selection of the optimal spectral regions (including removal of redundant variables) and the application of the optimal spectral pre-processing techniques[8] (i.e. 1st or 2nd derivatives, Savitzky-Golay smoothing, multiplicative scatter correction) or combinations of them are usually key steps in the multivariate modelling. They require a deep knowledge of the

sample under monitoring in order to avoid unrealistic results or loss of spectroscopic information.

In the food industry, when using vibrational spectroscopy for process control, it is common that the phenomenon we are interested in is hidden by other sources of variation from the main process[9]. In the case of 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 some organic acids that might be detrimental for the wine, as it is the case for acetic acid.

According to literature, in this type of situations ANOVA Simultaneous Component Analysis (ASCA) can be a very helpful tool to unravel the minor sources of variability. Amigo *et al.* studied the contribution of the three major design parameters of the bread staling process (enzyme treatment, measurement zone and storage time) by applying ASCA to NIR spectral data. They found that three effects were significant: the main one, 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[11]. 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 was independently evaluated. They found that time had always a significant effect in all intervals. The temperature and yeast strain factors showed a significant influence (p-value < 0.01) in some of the i-ASCA intervals, unlike classical ASCA results, which did not show significance for these factors[12].

The aim of this paper is to evaluate different methodologies to build ASCA models to study the variation associated to the wine alcoholic fermentation having a subprocess, which is an acetic acid bacteria (AAB) spoilage. We used different strategies to build the ASCA models, focusing on specific regions for certain factors or applying specific pre-processings to enhance the signals of minor compounds. Finally, different matrix unfolding procedures allowed us to study the individual factors of interest.

Materials and Methods

Fermentation

As in previous studies[13], grape must was obtained by dilution of concentrated white grape must (Mostos Españoles S.A., Ciudad Real, Spain). Dilution was performed with MilliQ quality water until a final sugar concentration of 200 $g \cdot L^{-1}$. Yeast assimilable nitrogen was adjusted using Actimaxbio* (Agrovin, Ciudad Real, Spain) and ENOVIT® (SPINDAL S.A.R.L. Gretz Armainvilliers, France) at a dosage of 0.3 $g \cdot L^{-1}$ for each additive to ensure proper fermentation performance.

Ten microvinifications were conducted in 500 mL conical flask with 350 mL each. Each microfermentation was inoculated with $3 \cdot 10^6$ $CFU \cdot mL^{-1}$ *Saccharomyces cerevisae* "E491" (Vitilevure Albaflor, YSEO, Danstar Ferment A.G., Denmark), following the instructions of the manufacturer to rehydrate dry yeast. Five microfermentations were intentionally deviated to simulate an acetic acid bacteria (AAB) contamination. Spoilage was simulated by adding *Acetobacter pasteurianus* grown in glucose medium (GY: 1% yeast extract; 1% glucose, w/v) (Cultimed, Barcelona, Spain). The strain was inoculated to reach a final concentration of $1 \cdot 10^6$ $CFU \cdot mL^{-1}$.

The fermentation process was kept under a constant temperature of 18 ºC until the end of alcoholic fermentation and it was monitored once a day by measuring sugars and acetic acid, using a Y15 analyzer (Biosystems, Barcelona, Spain). Alcoholic fermentation was considered finished when the sugar concentration was under Y15 analyzer limits of detection (LOD < 0.05 $g \cdot L^{-1}$).

Spectroscopic measurements

The analyses were carried out by taking 2 mL of each sample, centrifuging them and placing a drop on the reader of the portable 4100 ExoScan FTIR (Fourier Transform Infrared) spectrometer (Agilent, California, USA), equipped with an interchangeable spherical Attenuated Total Reflectance (ATR) sampling interface with a diamond crystal window. The spectroscopic range was from 4000 to 850 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. Spectra were collected using the Microlab PC software (Agilent,

California, USA) and data were saved as .spc files. The mean of the triplicates was used in subsequent data analysis.

The microvinifications – five fermentations under Normal Operation Conditions (NOC) and five acetic contaminated (AC) fermentations (acetic acid bacteria spoilage) – were analysed at eight sampling times. Spectra were arranged in a data matrix with dimensions 10x845x8 (IxJxK: samples x wavelengths x time points)

ANOVA-Simultaneous Component Analysis (ASCA)

ASCA aims at evaluating the significance of one or more experimental factors. It can be considered a direct generalization of analysis of variance (ANOVA) to multivariate data[14]. In this study, two known experimental factors are studied: the fermentation process and the contamination by AAB. Therefore, ASCA decomposes the original (centered) data matrix (X_c) according to:

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}\mathbf{m}^T = \mathbf{X}_{Fermentation} + \mathbf{X}_{Contamination} + \mathbf{X}_{Interaction} + \mathbf{E} \qquad \text{Equation 1}$$

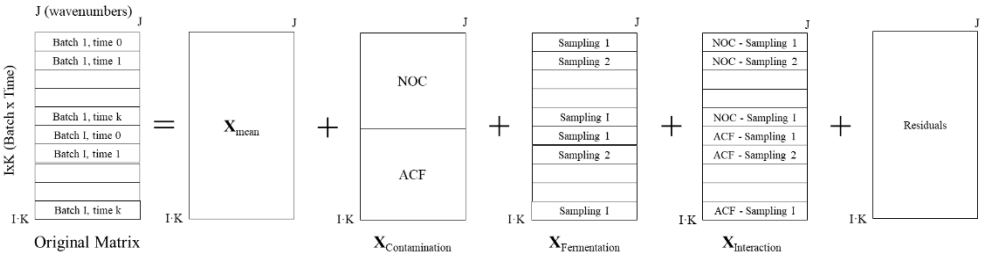


Figure 1. Scheme of the procedure applied to build IKxJ ASCA models.

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, $\mathbf{X}_{Fermentation}$ and $\mathbf{X}_{Contamination}$ are the matrices representing the effects of each one of the experimental factors, $\mathbf{X}_{Interaction}$ contains the interaction between 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. Thus, for example, "Contamination" factor has two levels of 40 observations each (5 batches per 8 sampling times), 40 observations will contain the average profile of NOC fermentations, and 40 will contain the average profile of AC fermentations[15]. The interaction matrix is calculated after the subtraction of the main effect matrices. Since all effect matrices are centered, the

magnitude of the effects can be evaluated as the sum of the squared matrix elements. Given a factor i:

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

where SSQ is the sum of squares of the elements in the matrix and $\|\cdot\|$ is the Euclidean norm. To evaluate if the effect of a particular factor or interaction is statistically significant, the SSQ value of the corresponding matrix is compared to its distribution under the null hypothesis (no experimental effect), as evaluated non-parametrically by a permutation test[16]. Given a factor i, and its associated matrix $\mathbf{X}_{\text{factor } i}$, the p-value is calculated as:

$$p - \text{value}(\mathbf{X}_i) = \frac{\text{nbr}(\mathbf{SSQ}(\mathbf{X}_{i,\text{perm}} \geq \mathbf{SSQ}(\mathbf{X}_{\text{factor } i}))) + 1}{k+1} \qquad \text{Equation 3}$$

where "nbr" is the number of occurrences, k is the number of permutations and $\mathbf{X}_{i,\text{perm}}$ is the matrix obtained after a random row permutation. Thus, the p-value indicates the number of cases where the variance of the studied factor is lower than the variance resulting from the permutation. In this way, the effect of the studied factor is compared to its distribution under the null hypothesis as estimated by the permutations[17,18]

Then, a bilinear decomposition of each effect 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 1 can be decomposed as:

$$\mathbf{X}_{\text{factor } i} = \mathbf{T}_{\text{factor } i} \cdot \mathbf{P}^{\mathrm{T}}_{\text{factor } i} + \mathbf{E}_{\text{factor } i} \qquad \text{Equation 4}$$

where $\mathbf{T}_{\text{factor } i}$ is the score matrix, $\mathbf{P}^{\mathrm{T}}{}_{\text{factor } i}$ is the loading matrix and $\mathbf{E}_{\text{factor } i}$ is the residual matrix of the i^{th} partitioned matrix in equation 1. The reduction of dimensionality enables a better visualization and interpretation of the data considering each experimental factor or interaction separately. The loadings for factor i define a subspace spanned by $\mathbf{X}_{\text{factor } i}$, that highlight the spectral directions related to the factor under study. The scores for factor i are the new coordinates of the observations on the Simultaneous Components (SCs) of the model[14].

ASCA models were built using Matlab R2021b (The MathWorks, Natick, USA) and PLS Toolbox v9.0 (Eigenvector Research Inc., Eaglerock, USA). Validation was performed using permutation tests and assessing the statistical significance (p-value), as implemented in PLS Toolbox[19]. The number of permutations was 10000[20] and results were considered to be statistically significant when p-value < 0.05. Throughout the article only the significant factors will be discussed.

Results and Discussion

Alcoholic fermentation and contamination evolution

Alcoholic fermentation of grape juice to get wine implies the transformation of about 200 $g \cdot L^{-1}$ of glucose and fructose into about 120 $mL \cdot L^{-1}$ of ethanol. In this study, all the fermentations took 190 hours to finish but, to better study the possible contaminations after this process, an additional sampling point (258 hours) was considered (Figure 2a). As all batches behaved similarly, they were all considered as fermentations under control, and thus their MIR spectra were used for further analysis. Acetic acid contamination consists of the biochemical oxidation of ethanol produced in the alcoholic fermentation into acetic acid[21]. However, it has to be pointed out that NOC fermentations also produce low amounts of acetic acid, since yeasts can synthetize this compound to obtain energy as part of their metabolism[22]. Thus, Figure 2b shows that, from 76 hours onwards, a different behaviour is observed in acetic acid production between NOC and AC fermentations, with a higher production of acetic acid for the contaminated process. At 258 hours, a final acetic acid concentration of 1.74 $g \cdot L^{-1}$ was reached, six times higher than for NOC.

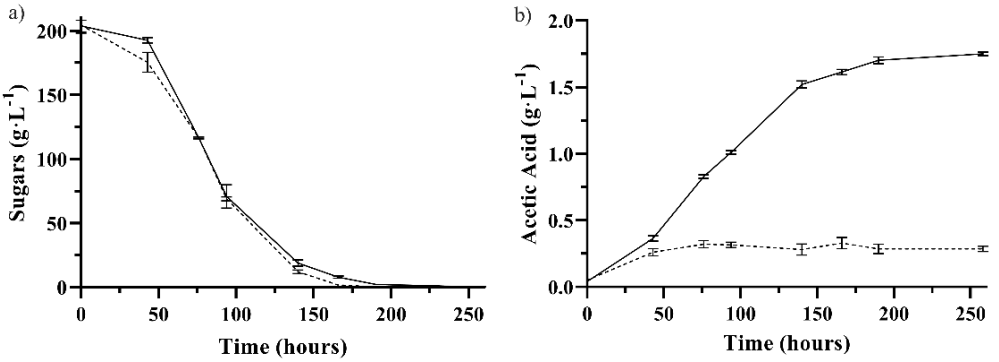


Figure 2. Evolution of (a) sugars and (b) acetic acid concentration during alcoholic fermentation. NOC fermentation (dashed line) and AC fermentation (continuous line).

The evolution of the ATR-MIR spectra during alcoholic fermentation and acetification is shown in Figure 3. As previously reported[9,23], the region showing the greatest variability during both processes is found between 950 and 1500 cm^{-1}. The absorption bands found in this region are related to CH_2, C–C–H, H–C–O bonds and C–C, C–O stretching vibrations. This region is considered the wine fingerprint region, since it is where the main components of the wine (sugars, acids, alcohols, phenolic compounds, etc.) show important absorption bands, although most of the differences found along the fermentation process are mainly related to changes in sugars and ethanol concentration[23].

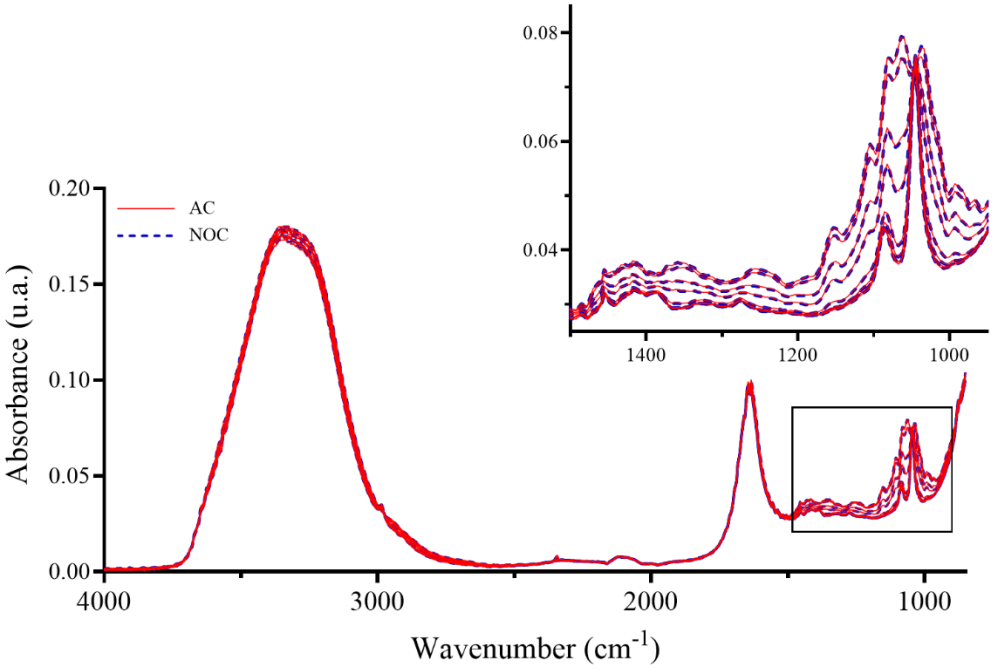


Figure 3. Evolution of raw ATR-MIR spectra during alcoholic fermentation and acetification. Upper spectra are a closer overlook of the most variable spectroscopic region. NOC fermentation (blue dashed line) and AC fermentation (red continuous line).

ASCA with IKxJ matrix unfolding

To study the variability of alcoholic fermentation when a subprocess occurs, in this case, a spoilage due to acetic acid bacteria, an ASCA model was built. ASCA requires the data to be arranged in a bidimensional matrix. For this model, an unfolding of the 3D matrix was performed along the row space (IKxJ), to have 80 spectra, which correspond to ten

batches (5 NOC and 5 AC) for 8 sampling times in the rows, and 845 wavelengths in the columns. Two factors were considered in this model: AAB contamination ("Contamination") and fermentation process ("Fermentation"), and their interaction. Different data pre-processing strategies were tested to envision the effect on the factors and similar results were obtained. The models built with raw and SNV pre-processed spectra (Table 1) will be discussed as examples.

Table 1. ASCA results for the 80x845 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 means the factor is significant.

Factors	**% Effect**	***p*-value**
Contamination	0.16	0.0001
Fermentation	98.95	0.0001
Contamination x Fermentation	0.11	0.1768
Residual	0.77	

The most relevant factor is the fermentation, accounting for 98.95% of the total variance. This is consistent with the fact that alcoholic fermentation comprises the main chemical changes regardless of whether NOC or AC are considered. Additionally, fermentation comprises other metabolism processes affecting minor compounds. Even though the contamination factor shows a low %Effect value (0.16%), it is significant. The contamination factor represents a subprocess, which occurs simultaneously with alcoholic fermentation and produces 1.74 $g \cdot L^{-1}$ of acetic acid. Every data pre-processing tested showed a similar value for each individual effect (Table S1).

Figure 4a shows the score values of the "contamination" submodel for each sample, coloured according to the type of process (NOC in green and AC in orange). Samples are grouped according to the type of process: NOC have negative values in SC1, while AC have positive ones. The first loading of the contamination factor (Figure 4b) shows that we can attribute this behaviour to the region between 1750 and 1000 cm^{-1}, not only to the regions where sugars absorb but also to the fingerprint region. It covers from 1500 to 1150 cm^{-1} and it is the result of absorption by proteins, acids, and many others molecules. The contribution of this region to the model here described may be directly related to acetic acid and other compounds involved in the metabolism of acetic acid bacteria[24].

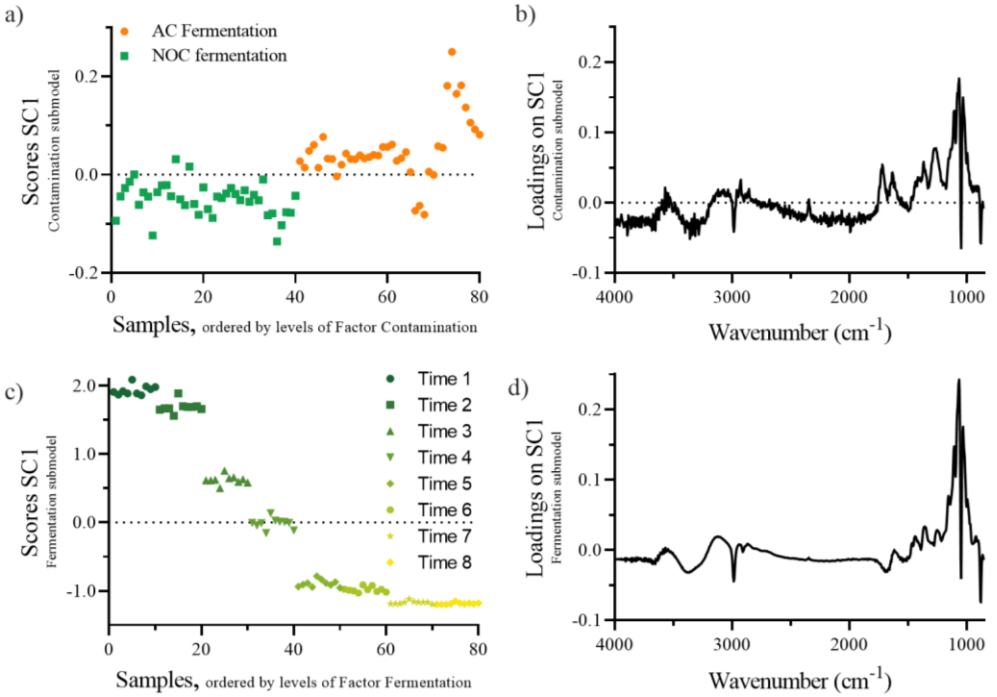
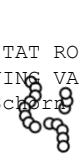


Figure 4. ASCA results on the IKxJ unfolded matrix (80 x 845). Scores of the first simultaneous component of the (a) contamination factor submodel and (c) fermentation factor submodel, where the samples are ordered by the levels of each factor. Loadings of the first simultaneous component of (b) contamination factor and (d) fermentation factor.

Concerning the factor "Fermentation", the evolution of the first score value (98,26% of variance) over time shows a sigmoidal trend (Figure S1). We have previously reported this behaviour for the first component in a PCA analysis of mid-infrared spectra6,7. The loading of this factor shows that the region between 1150 and 1000 cm^{-1} is the most related to this factor. Literature reports absorptions in this region related to ethanol and sugars in wine alcoholic fermentation and explained by the stretching modes of C–C and C–O bonds[13,23,24].

Grassi *et al.*[12] introduced the concept of interval ASCA (i-ASCA) to study the factors affecting specific regions of the FT–IR spectra. Typical FT–IR spectra can be described as large and small peaks, and even though the peaks are normalized with pre-processing methodologies, variability of the most intense peaks is still found as the most important. i-ASCA overcomes this problem by calculating ASCA models at different regions of the spectra. In our case, only the region between was selected, corresponding to the part of the



loading with more significance for the "Contamination" factor, after removing the region of the "Fermentation" factor, from 1157 to 977 cm^{-1}.

For this model, a IKxJ unfolding was also performed, as to have the spectra for each batch and each sampling time in the rows, and 171 wavelengths in the columns. The ASCA results are summarized in Table 2.

Table 2. ASCA results for the reduced 80x171 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 means the factor is significant.

	Raw		**SNV**	
Factors	**% Effect**	***p*-value**	**% Effect**	***p*-value**
Contamination	0.55	0.0001	0.25	0.0001
Fermentation	98.77	0.0001	98.90	0.0001
Contamination x Fermentation	0.16	0.0003	0.17	0.0003
Residual	0.52		0.68	

Despite the fermentation factor is still the main variability source, the contamination factor has gained importance compared to results in Table 1. For both factors, the score values show a similar trend and grouping (Table S2). ASCA results for a specific region agree with our previous research, in the sense that even when focussing on a specific region of acids, the alcoholic fermentation remains as the main factor[9]. This is because sugars also have major bands in the same region as acids[24].

Time interval ASCA

To deepen in the study of the variability when the AC subprocess occurs, the fermentation was divided into different parts taking into account the stages through which this process takes place. Specifically, four phases were considered: 1) stationary phase (yeast adaptation to the media and cellular growth); 2) tumultuous fermentation (maximum speed of the process is achieved); 3) tumultuous fermentation end (change in slope as deceleration occurs); and 4) process end (final sugar consumption and yeast death). Each phase of the process is represented by two sampling points. For this model, a IKxJ unfolding was performed, obtaining four data matrices with 20 spectra (five NOC and five AC at two sampling times) in rows and 845 wavelengths – or 171 wavelengths when using

the selected region – in columns. ASCA models were calculated for the four different parts of the process and the results are shown in Figure 5.

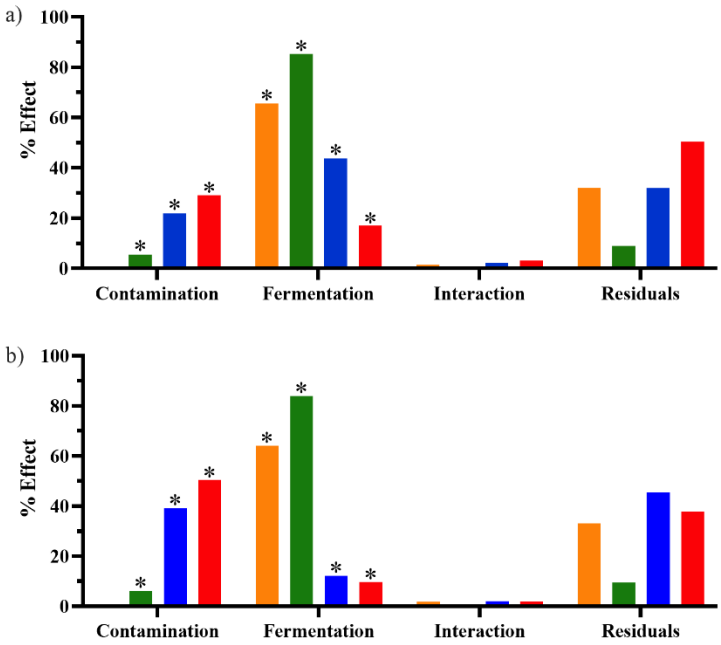


Figure 5. ASCA results of the different parts of the process. (a) Four matrices with dimensions 20x845 and (b) four matrices with dimensions 20x171. Orange: stationary phase, Green: tumultuous fermentation, Blue: tumultuous fermentation end, Red: process end. *: Significant factor (p-value < 0.05).

Figure 5a shows the % Effect for the factors "Contamination" and "Fermentation", divided into four phases of the process, using the whole spectra. The "Fermentation" factor shows most of the variability in every phase of the process, being significant in all cases. The highest value of variability (85.29% effect) for this factor is reached in the tumultuous fermentation, which is the part of the process with the maximum sugar transformation rate[22]. Moreover, in the first three phases of the process, the loadings for the "Fermentation" factor (Figure 6) show the interval between 1150 and 1000 cm^{-1} (associated to sugars and ethanol) as the main spectroscopic region correlated to this factor. The last phase of the fermentation process shows a noisy loading vector (Figure 6). In this phase, poor information of the process is obtained because the two sampling points collected coincide with the total depletion of sugars, as seen in Figure 2a.

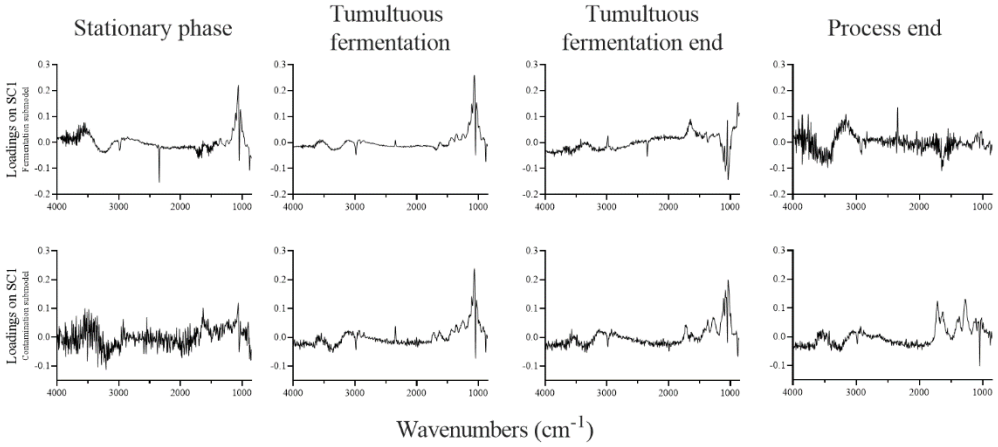


Figure 6. Loading of SC1 for fermentation factor (upper row) and contamination factor (lower row) submodels of the ASCA model. Each column represents a phase of the process.

Regarding the "Contamination" factor, it is significant from the tumultuous fermentation until the end of the process because acetic acid production increases from 76 hour (Figure 2b). However, the loading plot of this factor shows that the spectroscopic region where sugars absorb is the most important in the tumultuous fermentation (Figure 6). This can be explained as acetic acid bacteria may metabolize sugars at the beginning of the process, since there is a low ethanol concentration in the medium[21]. For the other phases, it can be observed that the fingerprint region increasingly gains importance, which may be attributed to the metabolism of acetic acid bacteria, especially acetic acid production. Carboxylic acid bonds show important absorbance bands, related to C=O stretching vibration at 1740 cm^{-1}, O–H bending and C-O stretching vibrations between 1200 and 900 cm^{-1} and C–H bending vibrations around 1400 and 1300 cm^{-1} [24].

Figure 5b shows the ASCA results obtained when using the selected region between 1795 and 1161 cm^{-1}. For the stationary and the tumultuous fermentation phases, a similar behaviour is found for both "Contamination" and "Fermentation" factors. However, once most of the sugars are transformed by yeast, the "Fermentation" factor decreases the % Effect value, and the "Contamination" factor gains importance. Thus, the "Contamination" factor is half of the variability (50.48% Effect) in the process end phase, since it reaches the highest acetic acid concentration, and the fermentation process has no effect because it has finished.

The results obtained demonstrate that time interval ASCA, with the whole spectra or a specific spectral region, allows focusing on a particular factor and its evolution. Additionally, for the other spectral pre-processings used, the contamination factor in the two last phases of the process (tumultuous fermentation end and process end) increases their % Effect reaching up to 70% of the variability when ASCA model is build using Savitzky-Golay smoothing or first derivative in combination to SNV pre-processed spectra (results reported in Figure S2).

The variance not explained by the ASCA model (residual term) could be explained by the biochemical differences between biological replicates of the same fermentation or acetification[25].

ASCA with IxJK matrix unfolding

To study the variability of the acetic acid bacteria spoilage, an ASCA model was built on a IxJK unfolded data matrix, with the spectra for each batch (five NOC and five AC) in the rows and 845 wavelengths for each one of the 8 sampling times (6760 variables) in the columns. In this model, only the "Contamination" factor was considered. Different pre-processing strategies were tested to study their effect on the factor. The ASCA model built using SNV pre-processed spectra (Figure 7) will be discussed as the model with minimum pre-processing. SNV was applied prior to matrix unfolding.

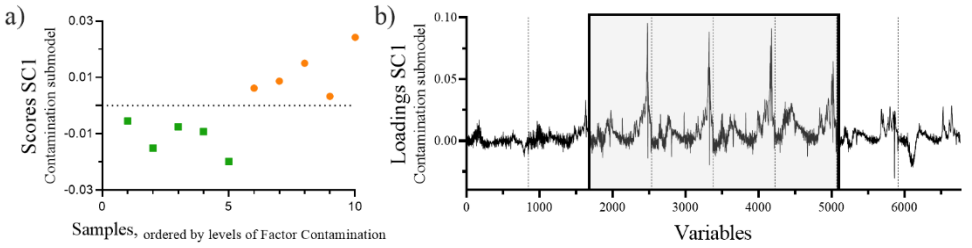


Figure 7. ASCA results on the IxJK unfolded matrix (10 x 6760) without pre-processing. Scores of the (a) contamination factor and (b) loading of the first simultaneous component of contamination factor. Vertical dotted lines divide each sampling point. Areas with larger peaks are highlighted in grey.

To better focus on the "Contamination" factor, every sampling point of a fermentation batch is unfolded in the same row, meaning eight consecutive spectra belonging to the same batch. The ASCA results show that the contamination factor accounts for 26.02% or

31.82% of the variability, for raw and SNV pre-processed spectra, respectively. Figure 7a shows the score values for the contamination factor. Different colouring stands for the type of sample: green for NOC and orange for AC. Samples are grouped according to the type of process: NOC have negative values in SC1, and AC have positive ones.

Figure 7b shows the loadings of the submodel for the contamination factor. It can be seen that the most important parts of the process are tumultuous fermentation and tumultuous fermentation end, as high absorption values are obtained from the third to the sixth sampling point and high loading values on SC1 from the 1690 to 5070 variables. Additionally, as it was stated along sections 3.2 and 3.3, both sugars and fingerprint regions are important parts of the spectra for the contamination factor.

Other pre-processing methodologies were tested in the IxJK unfolding ASCA model, and similar results were obtained. Sightly increases in the % Effect for contamination were achieved using smoothing and 1st derivative (Table S3) as this pre-processing amplify smaller peaks, such as the expected for acetic acid in the considered concentration.

Conclusions

To the best of our knowledge, this study shows for the first time the use of ASCA to study the variability of a wine fermentation evolution, such as the wine alcoholic fermentation process. Additionally, an intentionally provoked contamination with acetic acid bacteria was also studied, as the most unwanted microorganism in wineries. After applying different methodologies to arrange the spectral data and to enhance the information obtained from ASCA models, the results confirm that a subprocess, in this case bacterial contamination, can be detected with ASCA, when full spectra or specific spectral regions are used, when the process is divided in parts and when various spectral pre-processing approaches are applied if required.

Funding

Grant PID2019-104269RR-C33 funded by MCIN/AEI/ 10.13039/501100011033. This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-García, D.).

Acknowledgments

The authors would like to thank the Wine Biotechnology group for providing the acetic acid bacteria used in this experiment.

References (Paper 5)

1. Cozzolino, D. Advantages, Opportunities, and Challenges of Vibrational Spectroscopy as Tool to Monitor Sustainable Food Systems. *Food Anal Methods* **2022**, 15, 1390–1396, doi:10.1007/s12161-021-02207-w.

2. 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, 303–313, doi:10.1016/j.jfoodeng.2007.12.022.

3. Bisson, L.F. Stuck and Sluggish Fermentations. *Am J Enol Vitic* **1999**, 50, 107–119.

4. U.S. Department of Health and Human Services Food and Drug Administration Guidance for Industry Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance; **2004**.

5. Urtubia, A.; Pérez-correa, J.R.; 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.

6. Lee, L.C.; Liong, C.-Y.; Jemain, A.A. A Contemporary Review on Data Preprocessing (DP) Practice Strategy in ATR-FTIR Spectrum. *Chemometr Intell Lab* **2017**, 163, 64–75, doi:10.1016/j.chemolab.2017.02.008.

7. Miller, C.E. Chemometrics in Process Analytical Technology (PAT). Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries: Second Edition **2010**, 353–438, doi:10.1002/9780470689592.ch12.

8. Szymańska, E. Modern Data Science for Analytical Chemical Data – A Comprehensive Review. *Anal Chim Acta* **2018**, 1028, 1–10, doi:10.1016/j.aca.2018.05.038.

9. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

10. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.

11. Amigo, J.M.; del Olmo, A.; Engelsen, M.M.; Lundkvist, H.; Engelsen, S.B. 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. *Food Chem* **2019**, 297, 124946, doi:10.1016/j.foodchem.2019.06.013.

12. Grassi, S.; Lyndgaard, C.B.; Rasmussen, M.A.; Amigo, J.M. Interval ANOVA Simultaneous Component Analysis (i-ASCA) Applied to Spectroscopic Data to Study the Effect of Fundamental Fermentation Variables in Beer Fermentation Metabolites. *Chemometr Intell Lab* 2017, 163, 86–93, doi:10.1016/j.chemolab.2017.02.010.

13. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Monitoring Wine Fermentation Deviations Using an ATR-MIR Spectrometer and MSPC Charts. *Chemometr Intell Lab* **2020**, 201, 104011, doi:10.1016/j.chemolab.2020.104011.

14. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

15. Jansen, J.J.; Hoefsloot, H.C.J.; Van Der Greef, J.; Timmerman, M.E.; Westerhuis, J.A.; Smilde, A.K. ASCA: Analysis of Multivariate Data Obtained from an Experimental Design. *J Chemom* **2005**, 19, 469–481, doi:10.1002/cem.952.

16. De Luca, S.; De Filippis, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; 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.

17. Anderson, M.; Ter Braak, C.; Anderson, M.J.; F Ter Braak, C.J. Permutation Tests for Multi-Factorial Analysis of Variance. *J Stat Comput Simul* **2003**, 73, 85–113, doi:10.1080/00949650215733.

18. Ryckewaert, M.; Gorretta, N.; Henriot, F.; Marini, F.; Roger, J.M. 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.

19. Zwanenburg, G.; Hoefsloot, H.C.J.; Westerhuis, J.A.; Jansen, J.J.; Smilde, A.K. ANOVA–Principal Component Analysis and ANOVA–Simultaneous Component Analysis: A Comparison. *J Chemom* **2011**, 25, 561–567, doi:10.1002/cem.1400.

20. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA Simultaneous Component Analysis: A Tutorial Review. *Anal Chim Acta X* **2020**, 6, 100061, doi:10.1016/j.acax.2020.100061.

21. Mas, A.; Torija, M.J.; del Carmen García-Parrilla, M.; Troncoso, A.M.; Chen, Y.M.; Suh, S.J. Acetic Acid Bacteria and the Production and Quality of Wine Vinegar. *The Scientific World Journal* **2014**, doi:10.1155/2014/394671.

22. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

23. 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.

24. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

25. Moura, J.C.M.S.; Bonine, C.A.V.; de Oliveira Fernandes Viana, J.; Dornelas, M.C.; Mazzafera, P. Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants. *J Integr Plant Biol* **2010**, 52, 360–376, doi:10.1111/j.1744-7909.2010.00892.x.

Supplementary Material

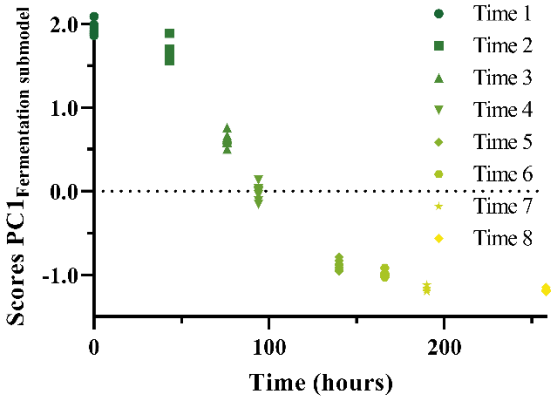
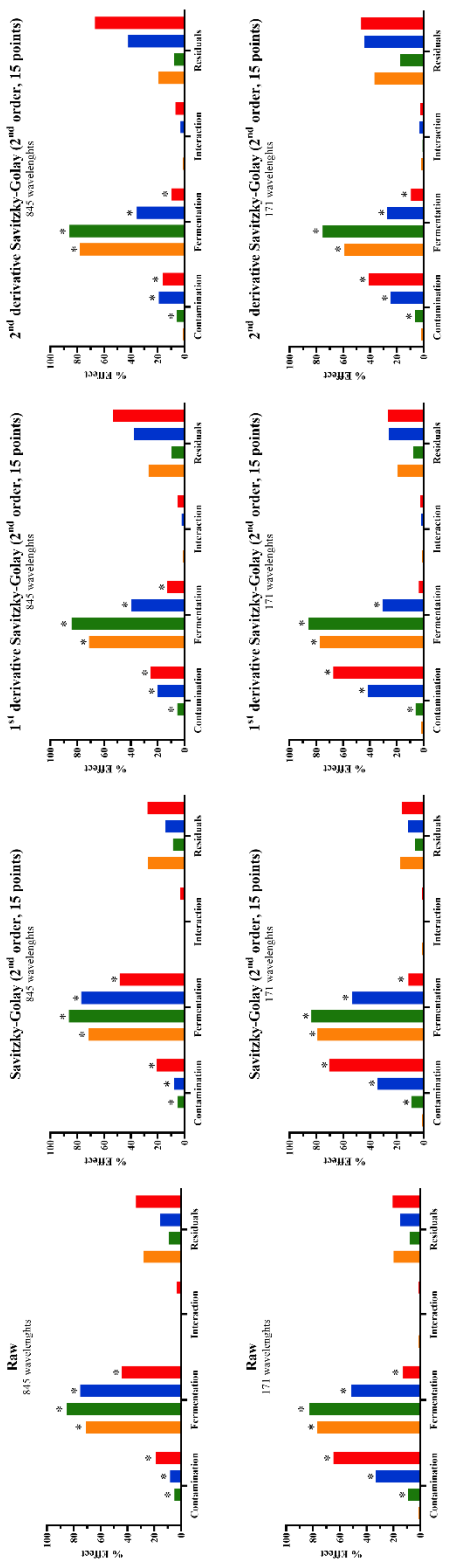


Figure S1. Scores of the first simultaneous component of the fermentation factor submodel evolution over time.



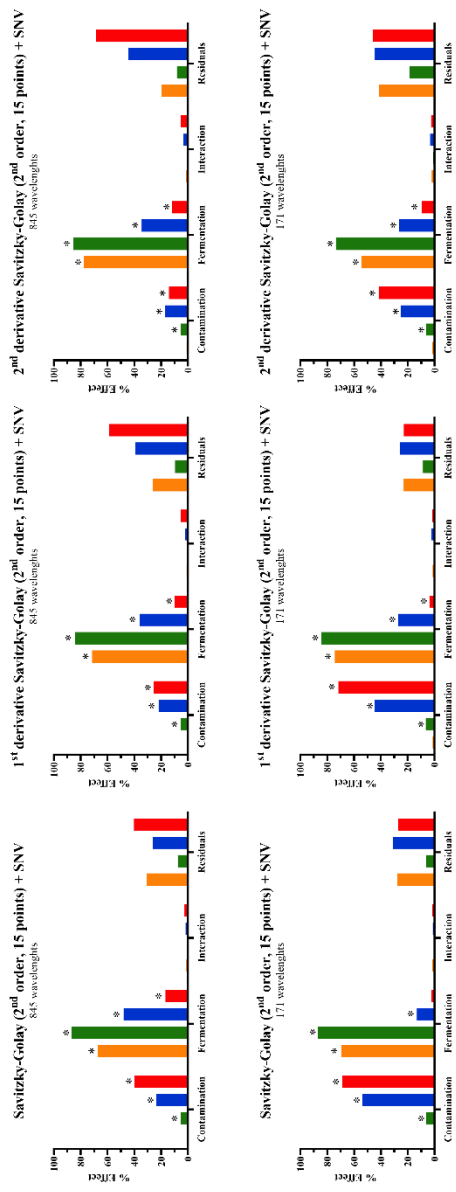


Figure S2. ASCA results of the different parts of the process using different pre-processings. Orange: stationary phase, Green: tumultuous fermentation, Blue: tumultuous fermentation end, Red: process end. *: Significant factor.

Table S1. ASCA results for the 80 x 845 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 reveals a significant factor.

	Savitsky-Golay (SG)		**1st Derivative**		**2nd Derivative**		**SG + SNV**		**1st Der SG + SNV**		**2nd Der SG + SNV**	
Factors	% Effect	p value	% Effect	p value	% Effect	p value	% Effect	p value	% Effect	p value	% Effect	p value
Contamination	0.16	0.0001	0.15	0.0001	0.15	0.0001	0.17	0.0001	0.16	0.0001	0.11	0.0001
Fermentation	99.02	0.0001	98.98	0.0001	99.12	0.0001	99.21	0.0001	98.97	0.0001	99.14	0.0001
Contamination x Fermentation	0.10	0.1938	0.16	0.0001	0.14	0.0001	0.11	0.0145	0.17	0.0001	0.15	0.0001
Residual	0.72	—	0.71	—	0.59	—	0.51	—	0.71	—	0.60	—

Table S2. ASCA results for the 80 x 171 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 reveals a significant factor

	Savitsky-Golay (SG)		**1st Derivative**		**2nd Derivative**		**SG + SNV**		**1st Der SG + SNV**		**2nd Der SG + SNV**	
Factors	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value
Contamination	0.55	0.0001	0.34	0.0001	0.36	0.0001	0.24	0.0001	0.41	0.0001	0.39	0.0001
Fermentation	98.88	0.0001	98.97	0.0001	97.98	0.0001	99.20	0.0001	98.80	0.0001	97.80	0.0001
Contamination x Fermentation	0.15	0.0001	0.20	0.0001	0.30	0.0001	0.15	0.0001	0.25	0.0001	0.34	0.0001
Residual	0.42	—	0.49	—	1.36	—	0.40	—	0.53	—	1.48	—

Table S3. ASCA results for the 10 x 6760 matrix, showing the percentage of variance (% Effect) for 'Contamination' factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 reveals a significant factor.

	SNV		**Savitsky-Golay (SG)**		**1st Derivative**		**2nd Derivative**		**SG + SNV**		**1st Der SG + SNV**		**2nd Der SG + SNV**	
Factors	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value	% Effect	*p* value
Contamination	31.82	0.0079	27.02	0.0001	30.20	0.0001	32.71	0.0001	35.70	0.0001	30.91	0.0001	32.77	0.0001
Residual	68.18	—	72.98	—	69.80	—	68.29	—	64.31	—	69.09	—	67.23	—

References (Section 4)

1. Du Toit, W.J.; Pretorius, I.S. The Occurrence, Control and Esoteric Effect of Acetic Acid Bacteria in Winemaking. *Ann Microbiol* **2002**, 52, 155–179.

2. Yamada, Y.; Yukphan, P. Genera and Species in Acetic Acid Bacteria. *Int J Food Microbiol* **2008**, 125, 15–24, doi:10.1016/j.ijfoodmicro.2007.11.077.

3. González, Á.; Hierro, N.; Poblet, M.; Mas, A.; Guillamón, J.M. Application of Molecular Methods to Demonstrate Species and Strain Evolution of Acetic Acid Bacteria Population during Wine Production. *Int J Food Microbiol* **2005**, 102, 295–304, doi:10.1016/j.ijfoodmicro.2004.11.020.

4. Guillamón, J.M.; Mas, A. Acetic Acid Bacteria. In *Biology of Microorganisms on Grapes, in Must and in Wine*; König, H., Unden, G., Fröhlich, J., Eds.; Springer Berlin Heidelberg, **2009**; pp. 31–46 ISBN 9783540854623.

5. Chai, Y.; Yang, H.; Zhao, L. Data Unfolding PCA Modelling and Monitoring of Multiphase Batch Processes. *IFAC Proceedings* **2013**, 46, 569–574, doi:10.3182/20130708-3-cn-2036.00058.

6. Golshan, M.; MacGregor, J.F.; Bruwer, M.J.; Mhaskar, P. Latent Variable Model Predictive Control (LV-MPC) for Trajectory Tracking in Batch Processes. *J Process Control* **2010**, 20, 538–550, doi:10.1016/j.jprocont.2010.01.007.

7. Camacho, J.; Picó, J. Online Monitoring of Batch Processes Using Multi-Phase Principal Component Analysis. *J Process Control* **2006**, 16, 1021–1035, doi:10.1016/j.jprocont.**2006**.07.005.

8. Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C. Breaking with Trends in Pre-Processing? *TRAC–Trends Anal Chem* **2013**, 50, 96–106, doi:10.1016/j.trac.2013.04.015.

9. Rinnan, Å.; Norgaard, L.; van den Berg, F.; Thygesen, J.; Bro, R.; Engelsen, S.B. Data Pre-Processing. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Elsevier Ltd, **2009**; pp. 29–50 ISBN 978-0-12-374136-3.

10. Rinnan, Å.; Berg, F. van den; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TRAC - Trends in Analytical Chemistry* **2009**, 28, 1201–1222, doi:10.1016/j.trac.2009.07.007.

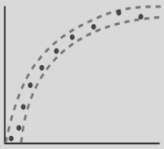
11. Oliveri, P.; Malegori, C.; Simonetti, R.; Casale, M. The Impact of Signal Pre-Processing on the Final Interpretation of Analytical Outcomes - A Tutorial. *Anal Chim Acta* **2018**, 1058, 9–17, doi:10.1016/j.aca.2018.10.055.

12. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.W.; Holmes, M.; Hanpin, M. Variables Selection Methods in Near-Infrared Spectroscopy. *Anal Chim Acta* **2010**, 667, 14–32, doi:10.1016/j.aca.2010.03.048.

13. Hoskuldsson, A. Variable and Subset Selection in PLS Regression. *Chemometr Intell Lab* **2001**, 55, 23–38, doi:10.1016/s0169-7439(00)00113-1.

14. Frans, S.D.; Harris, J.M. Selection of Analytical Wavelengths for Multicomponent Spectrophotometric Determinations. *Anal Chem* **1985**, 57, 2680–2684, doi:10.1021/ac00290a055.

15. Namkung, H.; Lee, Y.; Chung, H. Improving Prediction Selectivity for On-Line Near-Infrared Monitoring of Components in Etchant Solution by Spectral Range Optimization. *Anal Chim Acta* **2008**, 606, 50–56, doi:10.1016/j.aca.2007.10.047.



Section 5

A new Index to detect process deviations using IR spectroscopy and chemometrics process tools

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Deviation of alcoholic fermentation process

Despite the advances in both oenological technology and in knowledge regarding alcoholic fermentation, stuck and sluggish fermentations still represent a significant problem faced by wineries. An alcoholic fermentation is considered stuck or sluggish when the rate of sugar consumption is stopped or is too low for practical purposes in terms of duration and resources management[1]. Figure 1 illustrates various of these scenarios. Winemakers have difficulty restarting stopped or slow fermentations because even when the restart procedure is successful and fermentation can be completed, wine quality has usually already been affected. Therefore, it is crucial to detect and pinpoint the contributing factors that can lead to problematic fermentations[2].

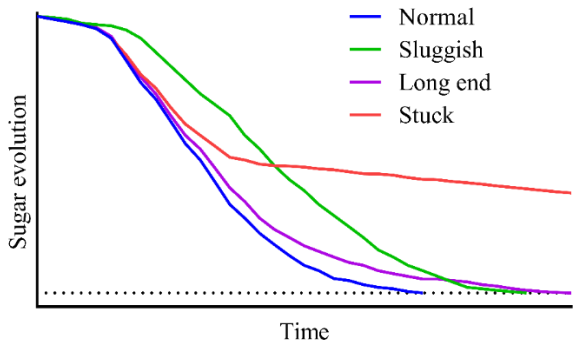


Figure 1. Types of problematic fermentations regarding sugar evolution over time. Adapted from Bisson[3].

Numerous factors have been identified as potential sources of sluggish and stuck fermentations. These include nutrient limitations, improper grape sanitary state, agricultural residues or improper temperature management, among others[4]. These factors generate stress conditions on yeast cells that directly affect their viability and population growth. This is because yeast metabolism suffers malfunctioning due to: the inability to synthesize essential compounds such as proteins; to the damage in yeast cellular integrity, as the lipid bilayer is sensitive to medium conditions; and to the incorporation of toxic nutrients into the cell[5–7].

In this chapter, the focus is on investigating sudden temperature changes and nitrogen deficiency as causes of sluggish fermentations. Temperature has a huge impact on yeast growth, especially when combined with the presence of ethanol, as both produce a synergic effect. The cell membrane is affected by the fermentation temperature in terms of less integrity and more permeability as the temperature increases. These conditions result in the incorporation of toxic compounds, such as ethanol, into the intracellular medium, as well as compounds that can disrupt the acid/base or redox balance[5,7].

Furthermore, nitrogen compounds that provides yeast assimilable nitrogen (YAN) to grape must are a key parameter for the wine fermentation process because this N is a substantial nutrient for yeast growth affecting the formation of yeast biomass. However, the optimal amount of YAN in must depends on various factors including yeast strain, the ratio of different nitrogen compounds and the initial sugar concentration[8,9]. If the nitrogen concentration is not enough, it is supplemented to the media through ammonia salts, amino acids and other minor compounds. These additions must be controlled since excessively high concentrations of YAN can cause problems such as the appearance of unpleasant odours related to the acceleration of fermentation with the subsequent stress on the yeasts[10].

In view of the problematic effects that both temperature and YAN may pose on the evolution of the fermentation process, it becomes essential to have simple and fast analytical methods available in the winery that can provide rapid responses at the early stages of deviation. Spectroscopic techniques, particularly MIR spectroscopy, offer a valuable tool of acquiring comprehensive chemical information from fermenting samples.

Multivariate control charts

The implementation of control charts built from high-dimensional data requires the application of dimensionality reduction techniques such as PCA or PLS. These methods are used to build models based on historical data and under normal operating conditions (NOC). PCA models focus on process variables, while PLS models incorporate both process variables and product quality data. Once the model is established, future data can be projected onto the model to determine if the process aligns with the expected patterns[11].

In this study, multivariate control charts based on PCA models were evaluated as several options can be found in the existing literature. These include control charts for each

independent principal component's score, control charts based on the Hotelling T^2 statistic, and control charts based on the Q-residuals of the PCA model. Each of these charts offers valuable insights into different aspects of the process, enabling effective monitoring and identification of deviations from the expected behaviour[11]. Figure 2 shows the results of these three examples of control charts for a fermentation that suffered a sudden temperature change (more information about the deviation in **Paper 6**). Scores are the projection of the samples in the new variable space. Hotelling T^2 is a statistical measurement of the distance of the samples from the centre of the model, and Q residuals represent the distance of the sample to the PCA space. Information of how to calculate them and how to establish statistical limits on Chapter 1.

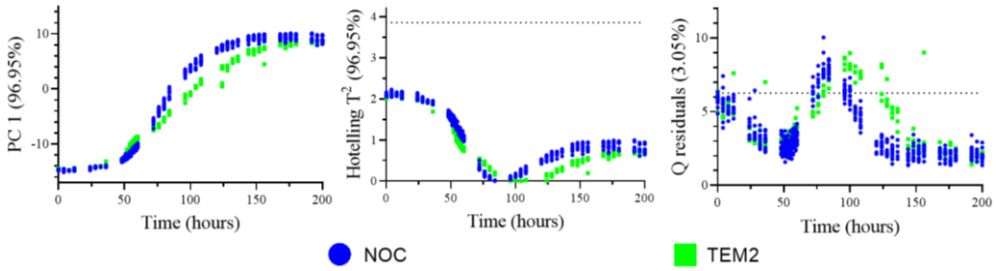


Figure 2. (a) Scores of the first principal component, (b) Hotelling T^2 and (c) Q residuals evolution over time. Normal Operating Conditions (NOC) fermentations in blue circles and fermentations that suffer a sudden temperature change in green squares. Dotted line show the statistical limit of each parameter.

The visual representation of the control charts allows for a comprehensive overview of the process performance, highlighting trends, patterns, and potential areas for improvement. However, even more detailed and specific information can be obtained by constructing additional control charts that incorporate combinations of these parameters as reported Bersimis *et al*.[12]. While most control charts in the literature are based on the PCA scores, to the best of our knowledge, no control chart based on the PCA loadings has been proposed. The loading matrix contains essential information about the significance of spectral bands or regions in the model[13]. This information not only contributes to the understanding of the model but also provides valuable chemical insights. By incorporating loading-based control charts, winemakers can gain a deeper understanding of the factors influencing the process and make more informed decisions.

Dissimilarity Index

The dissimilarity index (DI) proposed by Muncan *et al.*[14] caught our attention due to its ability to visualize the different stages of yogurt fermentation and its correlation with chemical information captured in the loadings. The DI employs a moving window approach, comparing spectra from a specific time frame with those from the immediately preceding time frame. The DI approach was applied to the alcoholic fermentation data but it did not allow to establish a control chart due to the high variability in the process rate at different stages. To solve this problem, an evolving window approach was proposed to compare spectra from the beginning of the process until a given moment with those from an initial time frame. This enabled us to effectively monitor alcoholic fermentation.

The monitoring of the alcoholic fermentation and the detection of deviations studied are detailed in **Paper 6**. These fermentation deviations were intentionally induced by suddenly increasing the temperature and promoting a nitrogen deficit that resulted in sluggish fermentation in both cases. When our new DI approach was applied to the data of these fermentations, the results demonstrate the efficacy of the evolving window dissimilarity index in monitoring wine alcoholic fermentation, identifying sluggish fermentation at early stages, and quantifying the impact of various factors on the process. In the best of our knowledge, this is the first control chart based on loading information, which can identify process deviation influencing process rate performance and obtain chemical information from the spectra.

A new Index to detect process deviations using IR spectroscopy and chemometrics process tools

Daniel Schorn–García[a], Jokin Ezenarro[a], Olga Busto[a], Laura Aceña[a], Ricard Boqué[a], Montserrat Mestres[a], Barbara Giussani[b]

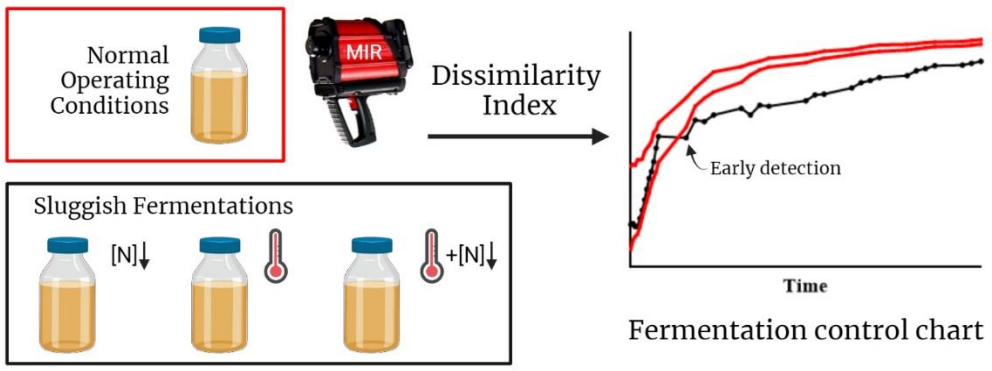
[a]Universitat Rovira i Virgili, Chemometrics and Sensorics for analytical solutions (ChemoSens) group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, 43007 Tarragona, Spain

[b] Dipartimento di Scienza e Alta Tecnologia, Università Degli Studi Dell'Insubria, 22100 Como, Italy

Submitted to Food and Biotechnology Process

Abstract

Process Analytical Technologies (PAT) have transformed the beverage production management by providing real-time monitoring and control of critical process parameters using non-destructive measurements, allowing for process readjustment if needed. New requirements call for new methods, and in this article, we propose a new method based on the construction of Multivariate Statistical Process Control (MSPC) charts from a new dissimilarity index capable of following fermentation processes. It has been applied to wine fermentation, the complex biochemical transformation of sugars into ethanol, which can be influenced by various factors. It allowed identifying deviations of the fermentation process in its early stages caused by nitrogen deficiency or temperature changes, and a combination of both. These results show the potential of this new approach to improve the monitoring and control of the key process stages, which allows maximizing quality and minimizing losses.



Keywords

mid-infrared spectroscopy; wine; real-time monitoring; evolving window principal component analysis; confidence limits

Introduction

The beverage industry is a continuously changing sector in which alcoholic drinks are the ones with the greatest economic contribution[1]. This implies that the alcoholic beverage industry today faces new challenges related to the different production processes, among which alcoholic fermentation stands out. This process is a complex biochemical transformation of sugars into ethanol that can slow down (sluggish fermentation) and even stop (stuck fermentation) for a number of reasons, including inadequate nutrients concentration or sudden changes in fermentation conditions[2,3]. The slow rate of sugar consumption by the yeasts indicates that their metabolism is not working properly, which could lead to the synthesis of undesirable molecules. These can negatively influence sensory perception of the final product and, therefore, its quality[4].

The possibility of recovering a problematic fermentation depends on early detection of the problem, which allows producers to take corrective measures to minimize the impact on product quality[5]. This concept perfectly fits with the implementation of Process Analytical Technologies (PAT). This approach is based on the idea that final product quality is achieved through real-time measurements throughout the process which allow readjustments while the problem is still solvable, rather than assessing product quality when the process has already been completed[6,7]. The implementation of PAT methodologies in the food and beverage industry requires in- or on-line to gather information about the molecules involved in the process. Additionally, the use of non-destructive measurements is recommended as they can provide valuable information without affecting the process' quality or composition[7]. Regarding wine alcoholic fermentation vibrational spectroscopy (near infrared, mid infrared or Raman) has proven to be a useful tool for obtaining process information[8–12]. In these cases, the data being acquired are multivariate and, therefore the use of Multivariate Statistical Process Control (MSPC) techniques is necessary for fermentation process monitoring and control. Among the different types of MSPC charts, those based on Principal Components Analysis (PCA) are widely used because they are simple and easy to interpret[13].

The application of the PCA-based control chart approach involves selecting a reference set that defines the normal operating conditions (NOC) for a particular process, with its intrinsic variability, and comparing future values with this set. Thus, a PCA is built using NOC data and future data are projected onto the 'NOC' PCA model. Then the values of the

statistical parameters Hotelling's T^2 and Q residual are used to assess whether the new data are compatible with those recorded in the NOC. Hotelling's T^2 and Q residuals control charts have been shown to be effective in wine alcoholic fermentation to detect lactic acid bacteria spoilage[8], along with other monitoring strategies[9,14]. Since PCA loadings capture the chemical information of the process, Muncan *et al*. propounded a Dissimilarity Index (DI) to compare loadings of a yogurt fermentation process using a moving window PCA approach[15]. In this way it was possible to distinguish and visualize three process stages related to the physicochemical properties of the yogurt matrix.

In this work, we present a Multivariate Statistical Process Control (MSPC) chart based on the Dissimilarity Index of NOC fermentations which, for the first time, was proposed to early detect process deviations. This improvement of the DI consists of using an evolving window approach which has been developed as a tool to monitor the progression of a biotechnological process such as wine fermentation. A portable Attenuated Total Reflectance-Fourier Transform Infrared (ATR−FTIR) spectrometer was used to perform at-line and non-destructive measurements, collecting spectra throughout the process. Three different scenarios that can lead to deviation in the fermentations were tested: sudden changes in temperature, nitrogen deficiency at the start of fermentation, and a combination of both.

Materials and Methods

Samples and fermentation characteristics

The samples were obtained from small-scale alcoholic fermentations (microfermentations) of grape must carried out under different conditions. Specifically, a commercial and concentrated white grape must (Juan Soler S.A., Cuenca, Spain) was diluted with MilliQ water to a final sugar concentration of 200 $g \cdot L^{-1}$ and, for each microfermentation, 1.5 L of diluted must was transferred into 2 L glass vessels for fermentation. Then, to start the alcoholic fermentation process, each must was inoculated with $3 \cdot 10^6$ $CFU \cdot mL^{-1}$ of *Saccharomyces cerevisae* Viniferm Revelación (Agrovin S.A., Ciudad Real, Spain), following the manufacturer instructions for rehydration of the dry yeast. Different fermentation conditions regarding the initial yeast assimilable nitrogen (YAN) concentration and temperature during the process were applied and these are summarized in Table 1 (together with the abbreviations used in the text).

Table 1. Experimental conditions for the different types of fermentation carried out.

Type of fermentation	Addition of N at the beginning	Temperature
NOC Normal Operation Condition	Yes	Constant at 20 ºC
NDef Nitrogen Deficit	No	Constant at 20 ºC
TEM1 Applied temperature gradient 1	Yes	Gradient (Fig. 1a)
TEM2 Applied temperature gradient 2	Yes	Gradient (Fig. 1b)
NDT Nitrogen Deficit and applied temperature gradient 2	No	Gradient (Fig. 1b)

YAN was adjusted for NOC, TEM1 and TEM2 fermentations using Actimaxbio* (Agrovin, Ciudad Real, Spain) and ENOVIT® (SPINDAL S.A.R.L. Gretz Armainvilliers, France) at a dosage of 0.3 $g \cdot L^{-1}$ each, to ensure proper fermentation performance. The initial YAN concentration was 120 $mg \cdot L^{-1}$, and after the adjustment, YAN increased to 232 $mg \cdot L^{-1}$.

NOC and NDef fermentations were kept at a constant temperature of 20 ºC throughout alcoholic fermentation. Two temperature gradients, showed in Figure 1, were applied to TEM1 and, TEM2 and NDT fermentations, respectively.

Two experimental plans were carried out with different number of samples and different conditions (diagram in Figure S1 and Table 2): the first consisting in 6 NOC, 4 NDef and 4 TEM1 fermentations; the second consisting in 6 NOC, 4 TEM2 and 4 NDT fermentations.

To follow the evolution of the different alcoholic fermentations, each of the samples were periodically monitored (exact times were specified in section 2.2) by measuring their density and pH, using a portable electronic densimeter (Densito2Go, Mettler Toledo, United States) and a portable pH-meter with a 201 T electrode (7+ series portable pH-meter, XS Instruments, Italy), respectively. Both instruments were calibrated daily using reference standards. Alcoholic fermentation was considered finished when density was less than 0.995 $g \cdot L^{-1}$. This value was reached between 7–9 days depending on the type of fermentation.

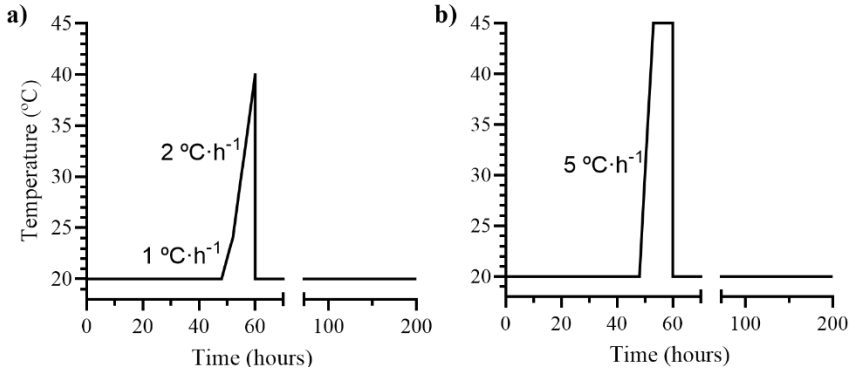


Figure 1. Temperature gradient applied to a) TEM1 fermentations and b) TEM2 and NDT fermentations.

Spectroscopic measurements

Infrared spectroscopic measurements were conducted according to a previously established methodology[5], using a portable 4100 ExoScan FTIR spectrometer (Agilent, California, USA) equipped with a spherical ATR sampling interface with a diamond crystal window. The spectroscopic range covered from 4000 to 850 cm^{-1}, and spectra were recorded at a resolution of 8 cm^{-1} and 32 scans, which imply the response measurements at 845 wavenumbers. To eliminate the interference due to variations in environmental conditions, an air-background spectrum was collected before each sample measurement. Each sample was analysed in triplicate by placing three drops of the sample on the crystal and recording the spectrum immediately after each drop. The spectra were collected using Microlab PC software (Agilent, California, USA) and saved as .spc files. The average of the three measurements was used for further data analysis. Spectra were collected at the same times of day, covering twelve hours per day of the process and taking samples every four hours. However, to better characterize the applied temperature gradient, one sample per hour was taken in NOC, TEM1, TEM2 and NDT. The details of the sampling points and matrix dimensions are shown in Table 2.

To maximize information and enhance small peaks, spectra were pre-processed using Savitzky-Golay second derivative (15 points and 2 order polynomial) and Standard Normal Variate.

Table 2. Overview of Replicates, Sampling Points and Data Matrix Dimensions.

Type of fermentation	**Number of replicates**	**Sampling points**	**3D matrix dimension** (replicates x samplings x wavenumbers)	**Unfolded matrix**
NOC	12	42	12 x 42 x 845	504 x 845
NDef	4	34	4 x 34 x 845	136 x 845
TEM1	4	42	4 x 42 x 845	168 x 845
TEM2	4	42	4 x 42 x 845	168 x 845
NDT	4	42	4 x 42 x 845	168 x 845

Dissimilarity Index

The dissimilarity index (Eq. 1) proposed by Muncan *et al*. is based on the absolute value of the product of the transposed first loading of each model and the first loading of the initial model[15].

$$A_i = 1 - \left|\mathbf{p}(i)^T \mathbf{p}(0)\right| \qquad \text{Equation 1}$$

where Ai is the dissimilarity index for model *i*, and $|\mathbf{p}(i)^T\ \mathbf{p}(0)|$ is the inner product (in absolute value) of the loading of a principal component for model i and a reference loading (initial model or model 0). This dimensionless parameter ranges from zero to one, with zero indicating that the loadings are identical and one indicating that they are orthogonal.

The DI was originally proposed using a moving-window approach to study a process, allowing to compare each part of the process with the preceding one in such a way that three process stages of the yogurt fermentation were distinguished.

Evolving Window Dissimilarity Index

The first step of the proposed approach involved the study of NOC fermentations. NOC alcoholic fermentations are monitored using evolving window principal component analysis (EWPCA) (Figure 2). This method calculates a PCA in increasing time intervals, allowing including the entire evolution of the process up to a specific sampling time and giving equal weight to each part of the process, in contrast to the moving window approach which only takes into account local process parts. The first loading of each model was

extracted and used to calculate the Evolving Window Dissimilarity Index (EWDI) along the fermentation process (Eq. 2).

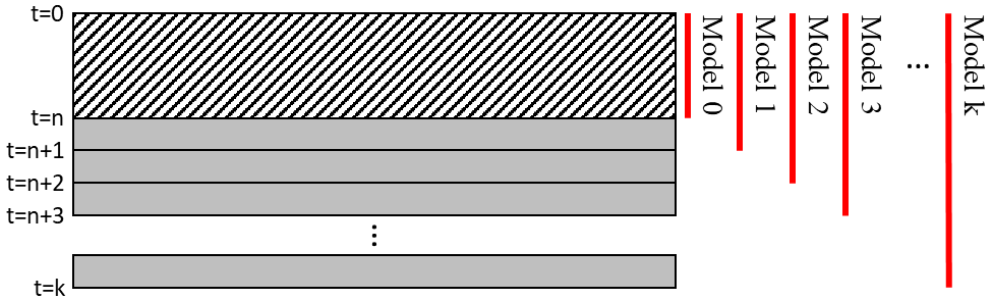


Figure 2. Scheme of the evolving approach used to calculate the evolving window dissimilarity index, from time 0 to time k.

$$DI_t = 1 - \left|(\mathbf{p}\,\mathrm{Model}_{0\to t})^T \cdot (\mathbf{p}\,\mathrm{Model}_{0\to n})\right| \qquad \text{Equation 2}$$

In the Figure 2, model 0 represents the start of fermentation, while model k represents the model encompassing the entire fermentation. The initial window (i.e., the time window of model 0) was optimized, using different initial time gaps (considered intervals, 0 to n, were between 0–24 hours and 0–36 hours). The loading extracted from model 0 was then used as the reference loading for NOC fermentations (p $\mathrm{Model}_{0\to n}$ in Equation 2) and the evolving window dissimilarity index was calculated for all NOC fermentations. NOC dissimilarity curves were used to build confidence limits calculated as the ± 2.201 standard deviations (Student's t for eleven degrees of freedom and 95%).

After optimization of model 0 for NOC fermentations, the EWDI was calculated for all the other fermentations conducted, i.e. fermentations carried out under deviations from normal conditions. The loading extracted from model 0 was also used as the reference loading (p $\mathrm{Model}_{0\to n}$ in Equation 2) for all other fermentations that were to be compared with NOC fermentations.

Analysis of Variance (ANOVA)-Simultaneous Component Analysis (ASCA)

Analysis of Variance (ANOVA)-Simultaneous Component Analysis (ASCA) was used to decompose the sources of variability influencing the data. ASCA is a multivariate

extension of ANOVA, which decomposes the variation in the data according to a predefined experimental design into the main effects and their binary combinations[16]. In this study, three variability factors were considered: (1) to which set of batches the fermentation belonged (experiment factor), (2) whether or not fermentation was subjected to a temperature gradient and (3) whether or not the fermentation had sufficient YAN at the beginning of alcoholic fermentation; and the interactions between them.

Results and Discussion

Optimization of the evolving window dissimilarity index

The optimization of the EWDI was performed by calculating the EWDI values of NOC spectra. The EWDI was calculated using the first PC loading, since in the PCA of alcoholic wine fermentation the first PC collects the main information of the process (see Fig. S2 and Schorn-García *et al.*[5]).The moving-window approach proposed to monitor yogurt fermentation[15] did not perform well in the case of wine alcoholic fermentation, as an erratic evolution (between 0 and 1) due to small differences in the process did not allow capturing all useful information about the process. Furthermore, a different initial number of sampling points included in the reference PCA (model 0) was tested (5 to 8 sampling points, which include spectra from 0 hours to 24 or 36 hours). An overview of the different EWDI evolution based on the number of sampling points included in model 0 is shown in Figure 3.

The EWDI evolution over time when considering 24 (Fig 3a) and 28 (Fig 3b) hours in model 0 was not useful due to a great increase in the EWDI value for the first sampling points, followed by a stabilization of the value. This behaviour did not allow studying the fermentation process, since after a certain point no differences were found between the EWDI values. However, the evolution over time of the EWDI considering 32 (Fig 3c) or 36 (Fig 3d) hours in the model 0 showed better results, probably due to the inclusion of the early stages of tumultuous fermentation (phase of the process with the highest sugar-to-ethanol transformation ratio). Thus, the 36-hour model 0 showed a better evolution in terms of sensitivity for the first sampling points, so this time was chosen to obtain the reference loading.

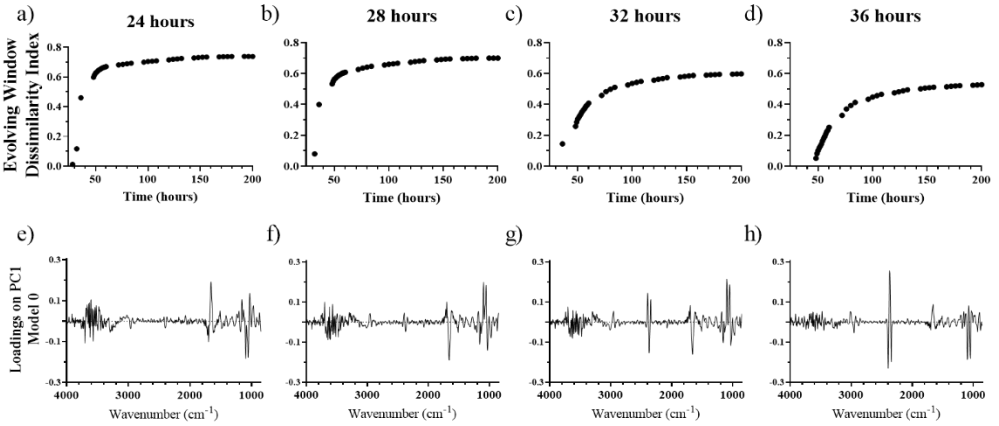


Figure 3. a-d) Evolving Window Dissimilarity Index (EWDI) based on loading variation of NOC wine alcoholic fermentation considering different time gaps for the reference loading (from 0 to 24 (a), 28 (b), 32 (c) or 36 (d) hours). X-axis indicates the last sampling point included in the evolving window. e-h) Reference loading used to calculate the EWDI, depending on the time gaps considered (from 0 to 24 (e), 28 (f), 32 (g) or 36 (h) hours).

Regarding the loading of model 0 (Fig 3e and f), it showed that the important regions were between 3000 and 3500 cm^{-1}, which could be attributed to the O−H stretching vibrations of water and ethanol, and between 1000 and 1800 cm^{-1}, corresponding to the fingerprint region of the alcoholic fermentation[11,17]. However, when more sampling points were considered (Fig 3g and h) in model 0, the calculated loading underweighted the region between 3000 and 3500 cm^{-1} and incorporated a signal from 2300 to 2400 cm^{-1}, which could be attributed to O=C=O stretching, as carbon dioxide is released during alcoholic fermentation[3].

Evolution of the process and deviations

The evolving window dissimilarity index of each sampling time was calculated for each fermentation type by grouping the spectra from each batch from the beginning to that sampling time and comparing the loadings of the first component with the loadings of the first component of model 0. For example, there were 12 NOC fermentations, so for hour 48 (first EWDI value in Figure 4) the loading of the PCA model obtained with a matrix of size 108x845 (9 sampling times x 12 microfermentations batches x 845 wavenumbers) was compared to the loading of model 0 of dimension 96x845 (8 sampling times x 12 microfermentations batches x 845 wavenumbers). The EWDI evolution over time is shown in Figure 4.

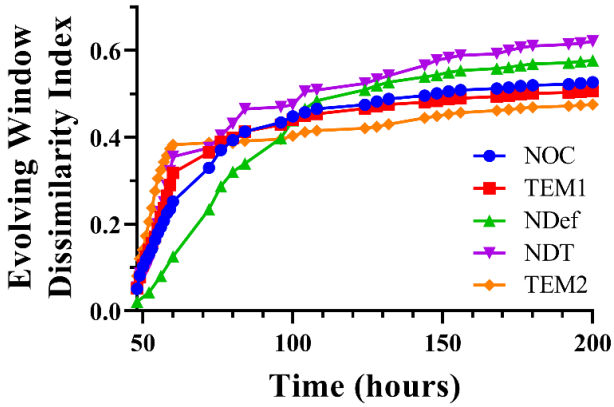


Figure 4. Evolving Window Dissimilarity Index based on loading variation of wine alcoholic fermentation with different characteristics (blue circles – Normal Operation Conditions (NOC); red squares – fermentations that suffered a temperature gradient depicted in Figure 1a (TEM1); green triangles – fermentations that started with a nitrogen deficit (NDef); purple inverted triangles – fermentations that started with a nitrogen deficit and suffered a temperature gradient depicted in Figure 1b (NDT); orange rhombi – fermentations that suffered a temperature gradient depicted in Figure 1b (TEM2)). X-axis indicates the last sampling point included in the evolving window.

In all types of fermentation, a noticeable increase of the EWDI was observed between 48 and 100 hours. This time frame coincided with the tumultuous phase of fermentation, during which the rate of glucose consumption was the highest. When sugars were converted to ethanol, EWDI levels also rose rapidly. After 100 hours, the increase in EWDI slowed down and continued until the end of fermentation, which occurred around 200 hours (when all types of fermentation finished). In general, the evolution of EWDI during fermentation provided valuable insights into the progress of the process, making it a useful tool for monitoring fermentation.

The fermentation types that underwent a temperature change (TEM1, TEM2 and NDT), showed a similar upward trend between 48 and 60 hours, when the temperature gradient was applied. This trend could be explained as yeasts have their optimal temperature range between 16 to 30 ºC[18]. At the beginning of the temperature gradient, the yeasts were in their optimal temperature range, which allows them to increase their metabolic activity. NDef showed a slower evolution over time; however, NDef and NDT ended the fermentation with a higher EWDI compared to NOC fermentation. The slower fermentation rate at the beginning would be explained by the lower amount of YAN available to carry out yeast metabolism properly[19].

Evolving Window Dissimilarity Index control chart

As discussed above, the NOC spectra of the first 36 hours was used to obtain the reference loading (first loading of model 0), calculated as the loading of the first PC (shown in Figure 4h). The NOC EWDI curves were used to calculate 95% confidence limits (average of 12 EWDI NOC curves ± 2.201 standard deviations) (red dotted lines in Figure 5).

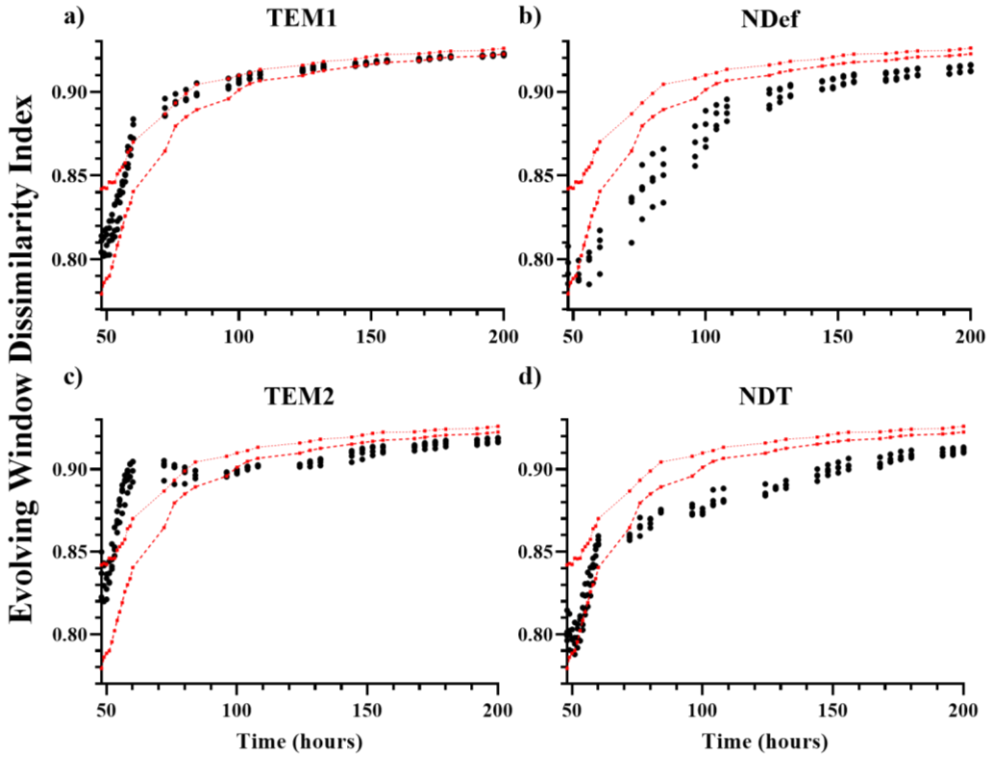


Figure 5. Evolving Window Dissimilarity Index (EWDI) control charts based on loading variation of wine alcoholic fermentation. Red dotted lines represent 95% confidence intervals calculated based on EWDI curves of the twelve NOC fermentations (average ± 2.201 standard deviation). The different charts represent the values of EWDI curves for different types of fermentation: a) fermentations that suffered a temperature gradient depicted in Figure 1a (TEM1); b) fermentations that started with a nitrogen deficit (NDef); c) fermentations that suffered a temperature gradient depicted in Figure 1b (TEM2); d) fermentations that started with a nitrogen deficit and suffered a temperature gradient depicted in Figure 1b (NDT). X-axis indicates the last sampling point included in the evolving window.

The NOC EWDI upper and lower control limits showed a greater difference up to 100 hours due to the great fermentation activity during this period. In fact, from 48 to 100 hours, the alcoholic fermentation process had the maximum rate of sugar consumption,

and even slight differences between NOC batches resulted in increased confidence limits. Then, from 100 hours to the end of the process, the confidence limits were much tighter as the fermentation process decreases in intensity.

However, despite the wider confidence limits at the beginning, the proposed MSPC chart was able to detect sluggish fermentations. Figure 5a shows the evolution of EWDI for TEM1, where the control chart was able to detect all the batches outside the confidence limit from 60 to 76 hours. In this type of sluggish fermentation, the EWDI showed out-of-control values shortly after the temperature rise reached its maximum at some later sampling points. However, from 80 hours onwards, TEM1 had a EWDI considered under control, and this behaviour can also be observed in Figure 4, where TEM1 had EWDI values similar to NOC at the end of the process. The time interval considered as out-of-control (60 to 76 hours) for TEM1 was relatively short, with only three sampling points being affected.

In the case of NDef (shown in Figure 5b), the control chart was able to detect all out-of-limits batches starting at 53 hours. This sluggish fermentation was intentionally induced by creating a nitrogen deficit, which can cause a malfunction in the metabolism of yeast [10]. Figure 5c shows the evolution of EWDI for TEM2. In this case, each batch was detected as out-of-control starting at 53 hours, with the exception of one particular time period where the EWDI values 'crossed' the confidence limit. The temperature variation applied between 48 to 60 hours was detected five hours after its initiation, indicating the effectiveness of the MSPC chart in detecting process deviations at an early stage. NDT fermentations (shown in Figure 5d) were a combination of nitrogen deficit and temperature gradient, resulting in intermediate behaviour compared to sluggish fermentations caused by each factor separately. The evolution of NDT EWDI showed that these fermentation batches were identified as out-of-control later than the other deviations studied, due to the slow progress caused by the nitrogen deficit, which was compensated by an increase in temperature that allowed an increase in yeast metabolism. However, after 72 hours, this type of sluggish fermentation remained out-of-control.

Study of the variability of sluggish fermentation

To study the influence of the different problems causing sluggish alcoholic fermentations and weight their contribution, an ASCA model was built. Since two groups of fermentations were performed (described in section 2.1.), this factor was also evaluated.

To build the ASCA model, all EWDI curves over time were used, i.e. 32 curves. Three factors were considered: the sudden change in temperature ("Temperature gradient" factor), the nitrogen deficit at the beginning of the fermentation (Nitrogen deficit" factor), the experiment group factor and their interaction. The results of the ASCA model are summarized in Table 3.

Table 3. ASCA results for the EWDI curves, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test. A *p*-value < 0.05 means the factor is significant.

Factor	**% Effect**	***p*-value**
Experiment	3.54	0.1286
Temperature gradient	15.76	0.0001
Nitrogen deficit	46.52	0.0001
Experiment x Temperature gradient	16.28	1.0000
Experiment x Nitrogen deficit	3.62	1.0000
Temperature gradient x Nitrogen deficit	8.45	1.0000
Residual	5.84	

First, it should be noted that the factor "experiment" and its combination with other factors did not show a significant effect, showing that both groups of fermentations could be compared. The most relevant factors were whether or not the fermentation was subjected to a temperature gradient and whether or not the fermentation started with the adequate nitrogen content. It should be noted that almost half of the variance (46.52%) was attributed to the "nitrogen deficit" factor. For this reason, it is necessary to guarantee the concentration of nitrogen (the minimum limit would depend on the concentration of sugar in the must) to ensure the end of fermentation and avoid the production of unwanted molecules[20,21]. These ASCA results highlighted the importance of measuring YAN to ensure a correct development of the alcoholic fermentation by yeast from the beginning of the process. On the other hand, the "temperature gradient" factor also showed a significant impact in the process evolution with a % effect close to 16%. This problem can be associated to inadequate temperature control, malfunctioning cooling jackets or high external temperatures which could lead to sudden rises in temperature. Furthermore, considering that alcoholic fermentation is an exothermic process, even in the absence of external factors, the process itself can cause the temperature to rise if not properly controlled. This can have a negative effect on the development of the fermentation process and on the

quality of the final product, due to the loss of aromas as a consequence of the increased volatilization of volatile compounds, as well as due to the stress of the yeasts, which could synthesize unwanted molecules[22,23].

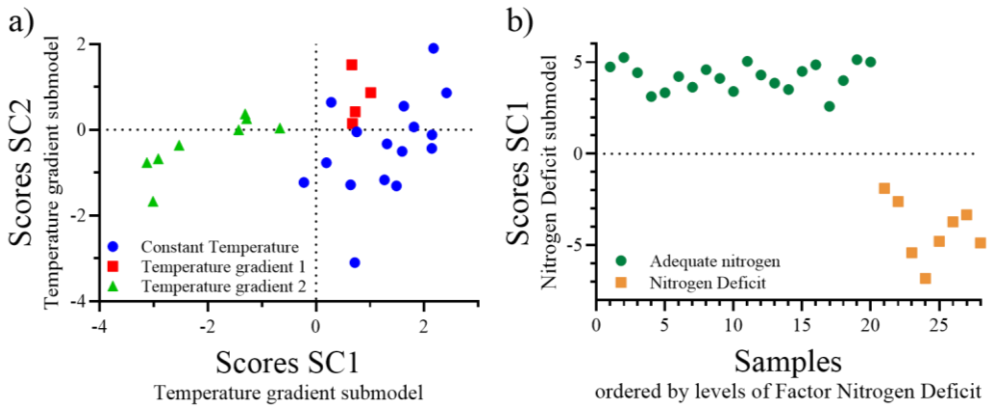


Figure 6. ASCA results on the EWDI curves. a) Scores of the first and second simultaneous component of the temperature gradient factor submodel and b) Score of the first simultaneous component of the nitrogen deficit factor submodel, where the samples are ordered by the levels of each factor.

ASCA results for each factor are depicted in Figure 6, which displays the scores of the principal simultaneous component. Figure 6a shows the score values of the "temperature gradient" submodel for each sample. The scores are coloured according to the temperature gradient applied to the process (blue circles for a constant temperature along the process, red squares for processes that suffered a temperature gradient as depicted in Figure 1a and green triangles for processes that suffered a temperature gradient as depicted in Figure 1b). Two main groups can be seen in the first simultaneous component: NOC, TEM1 and NDef have positive values in this component, while TEM2 and NDT have negative values. This could be explained because TEM1 showed, based on the results of sections 3.2 and 3.3, similar behaviour to NOC in terms of EWDI evolution, and that the second temperature gradient influenced the course of alcoholic fermentation and therefore in the evolution of EWDI over time.

Figure 6b shows the score value of the “nitrogen deficit” submodel for each sample, coloured according to whether or not the process had a nitrogen deficit at the beginning of fermentation (an adequate nitrogen content in green and a deficit of nitrogen in orange). In the first simultaneous component, two main groups can be observed. NOC, TEM1 and

TEM2 fermentations had positive values, while the NDef and NDT fermentations had negative values. The % effect of nitrogen deficit revealed by ASCA highlighted the importance of an adequate nitrogen content and resulted in a clear separation of score values between the groups.

Conclusions

The Dissimilarity Index (DI) is a process control parameter that can accurately describe the progression of fermentation in real-time. In this study, a modified DI was proposed and used to examine sluggish alcoholic fermentations and observe their development. The results have shown that the EWDI can be used to create a control chart based on data from normal operating conditions. Furthermore, the confidence limits of the control chart can be used to detect sluggish fermentations at an early stage. This is particularly important because early detection allows timely intervention to resolve any issues that may arise and therefore increases the chances of maintaining product quality.

In conclusion, the EWDI appears to be an effective tool for process control, allowing real-time monitoring and early detection of potential deviations. Additionally, the effect of different unwanted conditions affecting process kinetics have been studied and decomposed showing that EWDI evolution over time could indicate which kind of deviation is taking place and the importance of nitrogen supplementation at the beginning of the process. Further research would be developed to investigate the recovery of deviation using EWDI control charts.

Funding

Grant PID2019-104269RR-C33 funded by MCIN/AEI/ 10.13039/501100011033. This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-Garcia, D.). Grant URV Martí i Franqués - Banco Santander (2021PMF-BS-12).

Acknowledgments

The authors would like to thank the Wine Biotechnology group for the yeast strain used in this experiment. The authors would like to thank Víctor Maixé for his design of the fermentation container cover. The authors would like to thank Candela Ruiz-de-Villa for helping to set up the beginning of each experiment. Graphical abstract made using BioRender.com.

References (Paper 5)

1. Food Drink, E. Data & Trends EU Food and Drink Industry; Brussels, Belgium, **2022**.

2. Hernández, G.; León, R.; Urtubia, A. Detection of Abnormal Processes of Wine Fermentation by Support Vector Machines. *Cluster Comput* **2016**, 19, 1219–1225, doi:10.1007/s10586-016-0594-5.

3. Urtubia, A.; Pérez-correa, J.R.; 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.

4. 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.

5. Schorn-García, D.; Cavaglia, J.; Giussani, B.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy as a Process Analytical Technology in Wine Alcoholic Fermentation – A Tutorial. *Microchem J* **2021**, 166, 106215, doi:10.1016/j.microc.2021.106215.

6. U.S. Department of Health and Human Services Food and Drug Administration Guidance for Industry Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance; **2004**.

7. Jenzsch, M.; Bell, C.; Buziol, S.; Kepert, F.; Wegele, H.; Hakemey, C. Trends in Process Analytical Technology: Present State in Bioprocessing. In *New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins*; Kiss, B., Gottschal, U., Pohlscheidt, M., Eds.; Springer International Publishing: Switzerland, **2018**; pp. 211–252.

8. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. Monitoring Wine Fermentation Deviations Using an ATR-MIR Spectrometer and MSPC Charts. *Chemometr Intell Lab* **2020**, 201, 104011, doi:10.1016/j.chemolab.2020.104011.

9. Cavaglia, J.; Schorn-García, D.; Giussani, B.; Ferré, J.; Busto, O.; Aceña, L.; Mestres, M.; Boqué, R. ATR-MIR Spectroscopy and Multivariate Analysis in Alcoholic Fermentation Monitoring and Lactic Acid Bacteria Spoilage Detection. *Food Control* **2020**, 109, 106947, doi:10.1016/j.foodcont.2019.106947.

10. Cozzolino, D. Advantages, Opportunities, and Challenges of Vibrational Spectroscopy as Tool to Monitor Sustainable Food Systems. *Food Anal Methods* **2022**, 15, 1390–1396, doi:10.1007/s12161-021-02207-w.

11. Cozzolino, D. Sample Presentation, Sources of Error and Future Perspectives on the Application of Vibrational Spectroscopy in the Wine Industry. *J Sci Food Agric* **2015**, 95, 861–868, doi:10.1002/jsfa.6733.

12. 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, 282–297, doi:10.1080/05704928.2015.1132721.

13. Nomikos, P.; MacGregor, J.F. Multi-Way Partial Least Squares in Monitoring Batch Processes. *Chemometr Intell Lab* **1995**, 30, 97–108, doi:10.1016/0169-7439(95)00043-7.

14. Cavaglia, J.; Giussani, B.; Mestres, M.; Puxeu, M.; Busto, O.; Ferré, J.; Boqué, R. Early Detection of Undesirable Deviations in Must Fermentation Using a Portable FTIR-ATR Instrument and Multivariate Analysis. *J Chemom* **2019**, 33, e3162, doi:10.1002/cem.3162.

15. Muncan, J.; Tei, K.; Tsenkova, R. Real-Time Monitoring of Yogurt Fermentation Process by Aquaphotomics near-Infrared Spectroscopy. *Sensors (Switzerland)* **2021**, 21, 1–18, doi:10.3390/s21010177.

16. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-Simultaneous Component Analysis (ASCA): A New Tool for Analyzing Designed Metabolomics Data. *Bioinformatics* **2005**, 21, 3043–3048, doi:10.1093/bioinformatics/bti476.

17. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

18. Ribereau-Gayon, P.; Dubourdieu, D.; Doneche, B.; Lonvaud, A. Handbook of Enology: The Microbiology of Wine and Vinifications: Second Edition; Wiley, **2006**; Vol. 1; ISBN 9780470010365.

19. Martínez-Moreno, R.; Morales, P.; Gonzalez, R.; Mas, A.; Beltran, G. Biomass Production and Alcoholic Fermentation Performance of *Saccharomyces cerevisiae* as a Function of Nitrogen Source. *FEMS Yeast Res* **2012**, 12, 477–485, doi:10.1111/j.1567-1364.2012.00802.x.

20. Lleixà, J.; Martín, V.; Giorello, F.; Portillo, M.C.; Carrau, F.; Beltran, G.; Mas, A. Analysis of the NCR Mechanisms in *Hanseniaspora vineae* and *Saccharomyces cerevisiae* during Winemaking. *Front Genet* 2019, 10, 747, doi:10.3389/fgene.2018.00747.

21. Torrea, D.; Varela, C.; Ugliano, M.; Ancin-Azpilicueta, C.; Leigh Francis, I.; Henschke, P.A. Comparison of Inorganic and Organic Nitrogen Supplementation of Grape Juice – Effect on

Volatile Composition and Aroma Profile of a Chardonnay Wine Fermented with *Saccharomyces cerevisiae* Yeast. *Food Chem* **2011**, 127, 1072–1083, doi:10.1016/j.foodchem.2011.01.092.

22. Torija, M.J.; Rozès, N.; Poblet, M.; Guillamón, J.M.; Mas, A. Effects of Fermentation Temperature on the Strain Population of *Saccharomyces cerevisiae*. *Int J Food Microbiol* **2003**, 80, 47–53, doi:10.1016/s0168-1605(02)00144-7.

23. Woo, J.M.; Yang, K.M.; Kim, S.U.; Blank, L.M.; Park, J.B. High Temperature Stimulates Acetic Acid Accumulation and Enhances the Growth Inhibition and Ethanol Production by *Saccharomyces cerevisiae* under Fermenting Conditions. *Appl Microbiol Biotechnol* **2014**, 98, 6085–6094, doi:10.1007/s00253-014-5691-x.

Supplementary Material

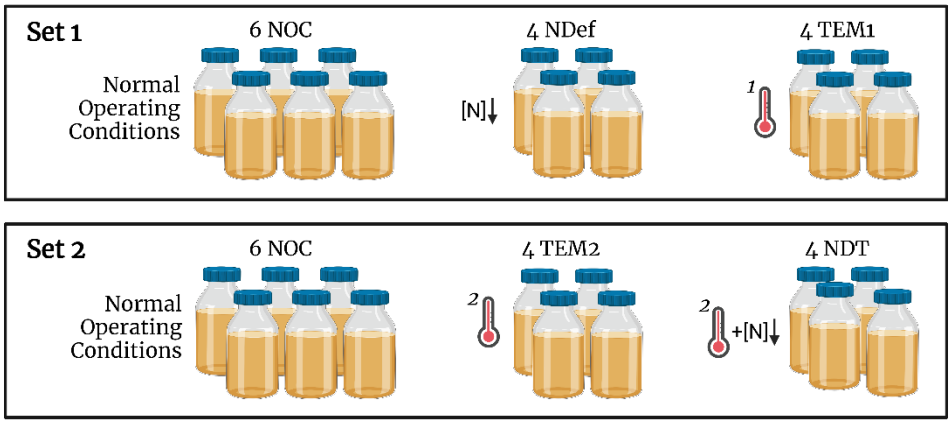


Figure S1. Schematic representation of the two set of alcoholic fermentation batches performed.

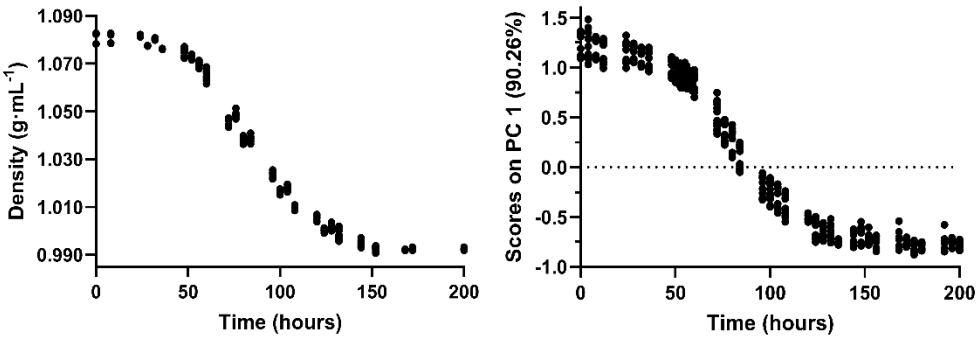


Figure S2. (left) Density and (right) PC1 score values evolution during NOC alcoholic fermentation.

References (Section 5)

1. Maisonnave, P.; Sanchez, I.; Moine, V.; Dequin, S.; Galeote, V. Stuck Fermentation: Development of a Synthetic Stuck Wine and Study of a Restart Procedure. *Int J Food Microbiol* **2013**, 163, 239–247, doi:10.1016/j.ijfoodmicro.2013.03.004.

2. Hernández, G.; León, R.; Urtubia, A. Detection of Abnormal Processes of Wine Fermentation by Support Vector Machines. *Cluster Comput* **2016**, 19, 1219–1225, doi:10.1007/s10586-016-0594-5.

3. Bisson, L.F. Diagnosis and Rectification of Arrested Fermentations. *Internet Journal of Viticulture and Enology* **2005**, 10.

4. Malherbe, S.; Bauer, F.F.; Du Toit, M. Understanding Problem Fermentations – A Review. *S Afr J Enol Vitic* **2007**, 28, 169–186, doi:10.21548/28-2-1471.

5. Valentine, G.D.S.; Walker, M.E.; Gardner, J.M.; Schmid, F.; Jiranek, V. Brief Temperature Extremes during Wine Fermentation: Effect on Yeast Viability and Fermentation Progress. *Aust J Grape Wine Res* **2019**, 25, 62–69, doi:10.1111/ajgw.12365.

6. Ivorra, C.; Pé Rez-Ortín, J.E.; Lí Del Olmo, M. An Inverse Correlation between Stress Resistance and Stuck Fermentations in Wine Yeasts. A Molecular Study. *Biotechnol Bioeng* **1999**, 64, 698–708, doi:10.1002/(sici)1097-0290(19990920)64:6.

7. Beltran, G.; Novo, M.; Guillamón, J.M.; Mas, A.; Rozès, N. Effect of Fermentation Temperature and Culture Media on the Yeast Lipid Composition and Wine Volatile Compounds. *Int J Food Microbiol* **2008**, 121, 169–177, doi:10.1016/j.ijfoodmicro.2007.11.030.

8. Beltran, G.; Esteve-Zarzoso, B.; Rozès, N.; Mas, A.; Guillamón, J.M. Influence of the Timing of Nitrogen Additions during Synthetic Grape Must Fermentations on Fermentation Kinetics and Nitrogen Consumption. *J Agric Food Chem* **2005**, 53, 996–1002, doi:10.1021/JF0487001.

9. Gutiérrez, A.; Chiva, R.; Sancho, M.; Beltran, G.; Arroyo-López, F.N.; Guillamón, J.M. Nitrogen Requirements of Commercial Wine Yeast Strains during Fermentation of a Synthetic Grape Must. *Food Microbiol* **2012**, 31, 25–32, doi:10.1016/j.fm.2012.02.012.

10. Martínez-Moreno, R.; Morales, P.; Gonzalez, R.; Mas, A.; Beltran, G. Biomass Production and Alcoholic Fermentation Performance of *Saccharomyces cerevisiae* as a Function of Nitrogen Source. *FEMS Yeast Res* **2012**, 12, 477–485, doi:10.1111/j.1567-1364.2012.00802.x.

11. Jackson, J.Edward. A User's Guide to Principal Components; Wiley, **1991**; ISBN 978-0-471-72532-9.

12. Bersimis, S.; Psarakis, S.; Panaretos, J. Multivariate Statistical Process Control Charts: An Overview. *Qual Reliab Eng Int* **2007**, 23, 517–543, doi:10.1002/QRE.829.

13. Kucheryavskiy, S. Elements of PCA Model. Accessed: 27th June 2023.

14. Muncan, J.; Tei, K.; Tsenkova, R. Real-Time Monitoring of Yogurt Fermentation Process by Aquaphotomics near-Infrared Spectroscopy. *Sensors (Switzerland)* **2021**, 21, 1–18, doi:10.3390/s21010177.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Chapter 4
General Discussion

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

The winemaking process encompasses a series of steps aimed at converting grapes into wines. These steps involve various chemical and biochemical transformations, contributing to the variability and complexity of the process. In this thesis, particular attention was given to grape maturity, alcoholic fermentation and crushed grape paste maceration, as they play an essential role in shaping the final product. To comprehensively monitor and investigate these processes, different spectroscopic techniques such as mid-infrared, near-infrared and Raman spectroscopy were used in combination with chemometric techniques.

Grape variability

Grapes, like any other food product, have been extensively studied by numerous authors regarding their inherent variability. This variability is not only influenced by year-to-year climate fluctuations, including the ongoing impact of climate change, but also by the specific field and vine characteristics. The field itself has been identified as a significant factor affecting grape characteristics. Various authors have reported variations in sugar concentration, acidity levels, and other constituents, not only between vines in the same field, but even between clusters of the same vine[1–4]. Such fluctuations can be attributed to many factors such as the characteristics of soil, the position of the vine within the vineyard, the distribution of metabolites within the vine or the unique biochemical signalling of each individual grape berry and also to interaction between them. These intricate interactions contribute to the diverse and complex nature of grape composition and highlight the need for careful control and management strategies to ensure consistent quality in viticulture.

To achieve precise control of grape composition and effectively manage variability, precision viticulture advocates for the use of tools that provide viticulturists with maximum information, preferably in a fast and non-destructive manner. Spectroscopy, a technique that meet both characteristics, has been proposed as a valuable tool for viticultural control[5]. However, applying spectroscopy in a complex process requires a thorough understanding of the parameters that influence it as well as the potential sources of variability. In this way it will be possible not only to relate spectrum bands or signals with specific chemical compounds and the existence of differences between grapes and vines but also to quantify the magnitude of their effect. This knowledge is essential for the development of viticultural practices aimed at more precisely controlling the

ripeness of the grape. This is because this accurate information about the factors that influence grape composition will provide a tool with which the viticulturists will be able implement targeted interventions and optimize their practices to achieve the desired results in terms of grape quality and vineyard management.

An extensive sampling was conducted in **Paper 2**, using an experimental design, to investigate the influence of positional effects on grape characteristics throughout the ripening process. The factors considered were the position of the grape bunch within the vine (divided into three positions from top to bottom) and the position of the grape grain within the bunch (divided into three positions along the conical geometry of the bunch, from top to bottom). When individual analyses were performed by measuring the traditional technological maturity assessment parameters, sugars (ºBrix) and acidity (pH), they showed minor differences between the different positional factors, but significant differences along ripening.

However, by using ANOVA-Simultaneous Component Analysis (ASCA), it became possible to study both chemical parameters simultaneously, considering that technological maturity is a ratio between sugars and acids. The most significant factor observed was the temporal evolution, as the sampling spanned five weeks. Additionally, the positional effect within the vine was also found to have a significant impact, which could be related to the different balance of metabolites import through the phloem[6].

In addition to the oenological parameters, grapes were analysed using MIR spectroscopy. This type of infrared spectroscopy includes the valuable fingerprint region (1800–900 cm^{-1})[7], where many chemical bonds absorb infrared radiation, thereby providing a specific signal for each sample. Initially, the spectra were used to predict the oenological parameters, with satisfactory results in terms of error and other performance parameters, confirming the potential of this technique for grape control[8,9].

Furthermore, the spectra obtained were used in conjunction with ASCA to study the sources of variability affecting grapes. The results objectively demonstrated the significant influence of all the considered positional effects during sampling, as well as their binary interactions. In addition, ASCA results showed the percentage of effects of the considered positional factors, showing that even individual berries are significantly different in the same bunch[10]. This outcome emphasises the ability of MIR spectroscopy not only to provide chemical information about the main constituents of grapes, such as

sugars and acids, but also to identify other minor compounds responsible for minor but significant differences. It should be noted that individual ASCA models were also built for each sampling time; however, due to the complexity of the evolving factors, although statistically significant, their interpretation was not straightforward[11].

Finally, the chemical information obtained from the spectra enabled the creation of a multivariate quality control chart based on the spectra from grapes considered optimal for winemaking. By projecting grapes from different sampling times onto the chart, we could observe that some grapes fell below the established limits. This result indicates that these grapes presented statistically similar profiles and were therefore suitable for vinification.

Viticultural and oenological variability

As previously stated, grapes introduce significant variability within the same field, even when considering the same cultivar harvested at the same time. This inherent variability poses a significant challenge for wineries, as they strive to produce wines with consistent quality and desired characteristics[12]. The task becomes even more complex if we take into account that wineries use many cultivars, so depending on the desired characteristics of the wine, they should harvest at different times[13].

In the case of red varieties, achieving optimal wine quality requires not only monitoring technological maturity but also phenolic maturity. Phenolic maturity encompasses aspects such as phenolic concentration and extractability, which greatly influence the colour, structure and overall sensory profile of red wines[14]. Variations due to phenolic maturity even within the same cultivar and harvest time adds another layer of complexity to the winemaking process.

Furthermore, wineries usually apply different oenological practices aimed at obtaining the desired colour extraction during winemaking. These practices include techniques such as extended skin contact with crushed grape prior to or after alcoholic fermentation, temperature control, the use of enzymes to facilitate phenolic extraction, and the manipulation of punch-down or pump-up mechanisms to optimize the extraction process[15–17]. Each of these practices introduces additional variability, making it even more challenging to monitor and account for all the sources of variability.

Shiraz, Merlot, and Cabernet Sauvignon varieties were selected for the study presented in **Paper 3**. The grapes were harvested at two different maturity states: one prior to optimal phenolic maturity (about 2 weeks before harvest) and the other at the optimal phenolic maturity state (harvest time). In addition, three different oenological practices were applied to explore phenolic extraction yield, varying in temperature and the use of enzymes and ranging from less to more extractive methods.

The wines obtained showed notable variations in the content of anthocyanins, total phenolic index, tannins, SO_2 resistant pigments, and colour density, with each parameter being influenced by the factors under consideration (grape variety, maturity and oenological practices). Merlot and Shiraz showed similar final characteristics, while Cabernet Sauvignon had higher values for all the parameters evaluated except for SO_2–resistant pigments. Moreover, wines harvested at the optimal maturity state, as expected, showed higher values for each parameter than unripe grapes. Regarding oenological practices, the intermediate extractive methods resulted in higher values of the parameters. It should be noted that excessive extraction can cause precipitation of certain compounds and trigger other reactions [18].

To investigate the impact of these factors (grape cultivar, maturity state and oenological practices) and how they interact, an ASCA model was built using the progress of the parameters during the winemaking process, which includes alcoholic fermentation and maceration time. The evolution of the ripening process explained a substantial portion of the observed variability, accounting for a third of the total. Both the grape cultivar and maturity state were found to have significant effects, underscoring the crucial role of the raw material in shaping the characteristics of the resulting wine[19]. Although the impact of oenological practices on the process and wine characteristics was comparatively lower, it was still deemed significant, indicating that winemakers have some degree of control over the process to achieve desired outcomes.

Furthermore, two spectroscopic techniques, NIR and MIR, were used to monitor the winemaking process and evaluate their efficacy. The results demonstrated that MIR spectroscopy outperformed NIR spectroscopy in terms of the amount of explained effects, with MIR accounting for 72.61% of the variability compared to 51.42% for NIR. It is worth noting that NIR spectroscopy, particularly in transmittance mode, is known to exhibit poorer performance compared to Attenuated Total Reflectance (ATR)–MIR spectroscopy[20]. Although MIR spectroscopy provided better insight into the evolution of the process over

time, it was less effective at capturing the variability arising from the factors considered. In contrast, NIR spectroscopy was more successful in capturing the grape related factors, although both techniques yielded significant results for the same factors and their binary combinations.

Both, **Paper 2** and **3** showed the importance of several grape factors and oenological practices in the characteristics of the grapes and wines obtained. These results highlight the need to fully understand the sources of variability to achieve the desired product. Moreover, as spectroscopy obtains information of the main bonds, this is the main molecules it would be also important to understand how these factors are affecting both products. Further research needs to be carried out, once the effect is quantified and factors organized in importance, to study how they affect chemical composition.

Raman spectroscopy

Even when external factors are considered, accounted for and controlled, alcoholic fermentation needs to be properly monitored. Close monitoring becomes imperative to prevent any negative impact on the process and ultimately on the quality of the wine[21]. However, many wineries rely on off-site analyses when a deviation is suspected, because specific analytical methods are often difficult to implement in the laboratories of small wineries. To avoid delays in obtaining results and enable corrective actions when necessary, the implementation of Process Analytical Techniques (PAT) is a good alternative[22].

The PAT approach involves closely monitoring the fermentation process to ensure wine quality from real-time measurements along the process. For this purpose, fast, robust and non-destructive techniques are highly desirable, and spectroscopy is often used for its advantages[23]. The implementation of spectroscopy can take different configurations depending on the placement of the spectrometer within the process. “At-line” approaches involve placing the spectrometers close to the process, “on-line” measurements divert the sample from the process for analysis, and “in-line” approaches use a sample probe directly inserted into the process.

Raman spectroscopy has proven to be very useful in analysing alcoholic fermentation[24], but its usefulness is further enhanced by the application of Spatially Offset Raman spectroscopy (SORS). SORS enables the analysis of fermentation through the container,

such as glass containers used in the experiment of **Paper 4**. In SORS, unlike conventional Raman measurements, the irradiation and the detection measurement of light are spatially displaced in such a way that a balanced subtraction of both allows obtaining spectral information of the container and of the sample inside[25].

To determine whether SORS was reliable and robust, two PAT approaches were tested in the experiment: "at-line" and "on-line". The at-line approach involved sampling the fermentation and analysing it at-line using a glass vial, while in the on-line approach the fermentation was analysed through the glass container in both the horizontal plane and the vertical plane from the bottom. Additionally, as carbon dioxide is produced during alcoholic fermentation and the gas bubbles push the yeast cells up, the potential interference of yeast in the spectra was also investigated. In the at-line configuration, the sample was analysed before and after being centrifuged to account for suspended yeast, and in the on-line approach the fermentation was analysed before and after shaken the container to promote the presence of suspended yeast.

A previously optimized methodology using Principal Component Analysis (PCA) to monitor alcoholic fermentation was employed to compare the different approaches. The results using one or two principal components demonstrated similar performance abilities for each configuration, except for the bottom measurement approach. This bottom approach was tested to understand whether the yeast deposited at the base of the container could provide information about the process through the Raman spectra. Additionally, to further test the monitoring capabilities, the prediction of oenological parameters such as density and pH was also assessed for each configuration. The best configuration for the prediction of both parameters was found to be with the spectra obtained through the horizontal plane, regardless of agitation. The prediction models obtained had similar performance to the ones in the bibliography[26].

Finally, to verify and quantify the effect related to the presence or absence of suspended yeast, an ASCA model was built for both approaches (the on–line approach before and after shaking and at–line approach before and after centrifugation). The results indicated that suspended yeast had no significant effect on the spectra regardless of the approach used. However, the at-line approach showed greater variability that may be explained due to the shorter pathway and some light scattering effects.

Considering the monitoring capabilities and the ability to predict oenological parameters, it was concluded that the on-line approach was superior for controlling alcoholic fermentation. This approach provided better signal quality in terms of fluorescence and ensured more representative spectra as the signal was collected directly from the process. Furthermore, as the suspended yeast had no impact on the final models, this technology could be applied in wineries using an on-line approach.

In **Papers 1** to **4**, it has been demonstrated that MIR, NIR and Raman spectroscopies are very useful analytical techniques in the grape and wine industry, and that every type of spectroscopic technique showed relevant information useful to obtain critical information of different parts of the process.

Acetic acid bacteria contamination

PAT approaches are valuable tools for monitoring the alcoholic fermentation process, as they enable a close and exhaustive examination of key parameters that influence this process, ensuring the overall quality of the final product. This is highly recommended, as even minor deviations in the fermentation process can have a pronounced impact on the organoleptic and physicochemical properties of the wine[27].

One of the most widespread and worrying concern in winemaking is the contamination caused by acetic acid bacteria (AAB), whose metabolism can lead to the formation of acetic acid. This acid is particularly troublesome due to its negative organoleptic impact in wines because, even at relatively low concentrations, this compound can impart unwanted vinegar-like flavours, significantly decreasing the quality and acceptance of the wine[28].

To address this challenge, further optimization of the PAT model is essential, with a specific emphasis on identifying the metabolic activities of AAB. Thus, whereas alcoholic fermentation process involves the conversion of sugars, typically around 200 grams per litre, into ethanol at a concentration of approximately 130 grams per litre, acetic acid formation must be carefully controlled to prevent it from exceeding the organoleptic threshold of 0.5 grams per litre. Although there is a significant difference in the order of magnitude of both processes, it is evident that both must be controlled due to the important impact that acetic acid formation has on the quality of the wine.

After monitoring both alcoholic fermentations and alcoholic fermentations contaminated with AAB with a MIR portable spectrometer, various data analysis techniques were used in **Paper 5** to enhance the information of this contamination in the spectra and increase the performance of future models. The use of ASCA made it possible to quantify the impact of this spoilage microorganism through different unfolding strategies, pre-processing techniques and targeted selection of process stages and spectral regions.

By using observation-wise unfolding, the significance of both alcoholic fermentation and acetification factors were found significant. However, the fermentation factor had an effect almost three orders of magnitude higher. The loading of the contamination factor and the alcoholic fermentation factor revealed a shared region in the spectra that explained both factors. To improve and direct the model towards the contamination factor, the shared region was excluded. An ASCA model was developed using only the spectral region ranging from 1795 to 1161 cm^{-1}, to focus on the contamination factor. In this spectroscopic region, acetic acid and other AAB metabolites show absorption bands[5]. However, despite removing the primary spectral region associated with sugars and ethanol, the fermentation factor remained the most significant, although the contamination factor gained more weight.

Considering the evolution of the scores over time in the fermentation submodel, four distinct process stages were identified and individually modelled. This approach determined a greater influence of the contamination factor, particularly after the main fermentation stage known as "tumultuous fermentation". This effect was further amplified when focusing on the selected spectral region. The tumultuous fermentation stage, characterized by a rapid sugar-to-ethanol conversion, hinders the contamination-related information, as major chemical changes occur in the fermenting matrix.

Additionally, batch-wise unfolding showed promising results in terms of emphasizing the contamination factor, while not considering the variability associated with the fermentation factor. Furthermore, different pre-processing techniques were tested in every approach, showing that spectral pre-processing can also enhance the impact of contamination, especially by accentuating small peaks through derivative transformations[29]. These findings further enriched the understanding of ABB contamination process and facilitated the development of more effective models for monitoring and control purposes.

The results of **Paper 5** highlight the importance of performed a proper data analysis methodology prior to modelling. This is of utmost importance, especially when subtle but important deviation occur in the process.

Multivariate Statistical Process Control

There is a growing recognition of the importance of early detection of process deviations, which allows the recovery of the fermentations as fast as possible in order to maintain the quality of the final product[30]. Vibrational spectroscopy has emerged as a valuable tool for monitoring alcoholic fermentation, as evidenced by the results presented in **Papers 1 to 5**. By using this analytical technique, it becomes possible to obtain multivariate data that can be used to establish robust process control approaches.

Regarding multivariate statistical process control, various approaches have been used to develop control charts based on data reduction algorithms such as Principal Component Analysis (PCA) or Partial Least Squares (PLS) regression. These methods typically rely on scores, Hotelling's T^2, or Q residuals to determine the statistical limits of these parameters calculated based on Normal Operating Conditions (NOC). Afterwards, the observed process data is compared against these predetermined limits, enabling the detection of deviation and the corrective actions to be taken[31,32].

However, despite the effectiveness of these control chart approaches, there has been a notable gap in the literature regarding the incorporation of information from the loadings of the model. Loadings contain valuable chemical information of the process and allow not only the detection but the interpretation of the cause of a potential deviation. To address this gap, in **Paper 6** an optimized version of the Dissimilarity Index initially proposed by Muncan *et al.*[33] is proposed. This novel approach enhances the ability to detect deviations by comparing the loading profiles of different process stages. It compares the loading for a PCA model from a time interval of the process to a reference loading obtained from a time interval of a previous stage of the process.

Particular attention was given to selecting the optimal time frame for the reference loading. Four distinct time frames, ranging from 24 to 36 hours, were considered to ensure that early deviations could be captured and meaningful process information could be obtained. After analysis of the loadings, it was observed that the most comprehensive chemical information related to the fermentation process was captured when using the

reference loading obtained from the model using spectra collected from the beginning until 36 hours.

To validate the efficacy of this approach, four different types of deviations were intentionally promoted and monitored by a MIR portable spectrometer. These deviations included two types of sudden temperature changes during the fermentation process, initial must nitrogen compound deficiencies and a combination of both. Using the calculated NOC dissimilarity index, the statistical limits were calculated using the *t*-student distribution at 95% of confidence level. Remarkably, the optimized approach successfully detected all deviations at early stages, and, importantly, three of the cases remained deviated throughout the fermentation process. These findings allowed us to determine that this deviation, which mainly affects the kinetics of the process, also had a large impact on other chemical information shown in the loadings. Yeast metabolism may produce metabolites as stress-response, which would impact the organoleptic properties of wines.

Additionally, to deepen the knowledge of both types of deviations (sudden temperature changes and initial nitrogen deficit), an ASCA model was built to quantify the impact of the deviation and assess which one had the greatest impact. The results showed that both deviations were statistically significant, but that a nitrogen deficit causes a greater impact on the process. This reinforces the need for carefully considering the initial fermentation conditions in the must to assure a good process performance.

Throughout the different **Papers**, it has been shown that vibrational spectroscopy is a valuable tool to control grape maturity, alcoholic fermentation and maceration. However, as highlighted in **Paper 6**, this analytical technique not only allows to monitor the fermentation but can also to be used to early detect deviations, which facilitates the application of corrective measures to avoid both product and economic losses.

References (General Discussion)

1. Shahood, R.; Torregrosa, L.; Savoi, S.; Romieu, C. First Quantitative Assessment of Growth, Sugar Accumulation and Malate Breakdown in a Single Ripening Berry. *OENO One* **2020**, 54, 1077–1092, doi:10.20870/oeno-one.2020.54.4.3787.

2. Robinson, S.P.; Davies, C. Molecular Biology of Grape Berry Ripening. *Aust J Grape Wine Res* **2000**, 6, 175–188, doi:10.1111/j.1755-0238.2000.tb00177.x.

3. Haselgrove, L.; Botting, D.; van Heeswijck, R.; Høj, P.B.; Dry, P.R.; Ford, C.; Iland, P.G. Canopy Microclimate and Berry Composition: The Effect of Bunch Exposure on the Phenolic Composition of *Vitis vinifera* L. cv. Shiraz Grape Berries. *Aust J Grape Wine Res* **2000**, 6, 141–149, doi:10.1111/j.1755-0238.2000.tb00173.x.

4. Kontoudakis, N.; Esteruelas, M.; Fort, F.; Canals, J.M.; De Freitas, V.; Zamora, F. Influence of the Heterogeneity of Grape Phenolic Maturity on Wine Composition and Quality. *Food Chem* **2011**, 124, 767–774, doi:10.1016/j.foodchem.2010.06.093.

5. Bureau, S.; Cozzolino, D.; Clark, C.J. Contributions of Fourier-Transform Mid Infrared (FT-MIR) Spectroscopy to the Study of Fruit and Vegetables: A Review. *Postharvest Biol Technol* **2019**, 148, 1–14, doi:10.1016/j.postharvbio.2018.10.003.

6. Coombe, B.G.; McCarthy, M.G. Dynamics of Grape Berry Growth and Physiology of Ripening. *Aust J Grape Wine Res* **2000**, 6, 131–135, doi:10.1111/j.1755-0238.2000.tb00171.x.

7. Cozzolino, D. Sample Presentation, Sources of Error and Future Perspectives on the Application of Vibrational Spectroscopy in the Wine Industry. *J Sci Food Agric* **2015**, 95, 861–868, doi:10.1002/jsfa.6733.

8. Musingarabwi, D.M.; Nieuwoudt, H.H.; Young, P.R.; Eyéghè-Bickong, H.A.; Vivier, M.A. A Rapid Qualitative and Quantitative Evaluation of Grape Berries at Various Stages of Development Using Fourier-Transform Infrared Spectroscopy and Multivariate Data Analysis. *Food Chem* **2016**, 190, 253–262, doi:10.1016/j.foodchem.2015.05.080.

9. 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.

10. Pagay, V.; Cheng, L. Variability in Berry Maturation of Concord and Cabernet Franc in a Cool Climate. *Am J Enol Vitic* **2010**, 61, 61–67.

11. Parpinello, G.P.; Nunziatini, G.; Rombolà, A.D.; Gottardi, F.; Versari, A. Relationship between Sensory and NIR Spectroscopy in Consumer Preference of Table Grape (cv Italia). *Postharvest Biol Technol* **2013**, 83, 47–53, doi:10.1016/j.postharvbio.2013.03.013.

12. Setford, P.C.; Jeffery, D.W.; Grbin, P.R.; Muhlack, R.A. Factors Affecting Extraction and Evolution of Phenolic Compounds during Red Wine Maceration and the Role of Process Modelling. *Trends Food Sci Technol* **2017**, 69, 106–117, doi:10.1016/j.tifs.2017.09.005.

13. Kader, A.A. Fruit Maturity, Ripening, and Quality Relationships. *Acta Hortic* **1999**, 485, 203–208, doi:10.17660/actahortic.1999.485.27.

14. Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Dubourdieu, D. Handbook of Enology, The Chemistry of Wine: Stabilization and Treatments: Second Edition; Wiley, **2006**; Vol. 2; ISBN 9780470010396.

15. Federico Casassa, L.; Beaver, C.W.; Mireles, M.S.; Harbertson, J.F. Effect of Extended Maceration and Ethanol Concentration on the Extraction and Evolution of Phenolics, Colour Components and Sensory Attributes of Merlot Wines. *Aust J Grape Wine Res* **2013**, 19, 25–39, doi:10.1111/ajgw.12009.

16. Pérez-Navarro, J.; García Romero, E.; Gómez-Alonso, S.; Izquierdo Cañas, P.M. Comparison between the Phenolic Composition of Petit Verdot Wines Elaborated at Different Maceration/Fermentation Temperatures. *Int J Food Prop* **2018**, 21, 996–1007, doi:10.1080/10942912.2018.1479856.

17. Pascual, O.; Ortiz, J.; Roel, M.; Kontoudakis, N.; Gil, M.; Gómez-Alonso, S.; García-Romero, E.; Canals, J.M.; Hermosín-Gutíerrez, I.; Zamora, F. Influence of Grape Maturity and Prefermentative Cluster Treatment of the Grenache Cultivar on Wine Composition and Quality. *OENO One* **2016**, 50, 169–181, doi:10.20870/oeno-one.2016.50.4.1824.

18. Kelebek, H.; Canbas, A.; Selli, S.; Saucier, C.; Jourdes, M.; Glories, Y. Influence of Different Maceration Times on the Anthocyanin Composition of Wines Made from *Vitis vinifera* L. cvs. Boğazkere and Öküzgözü. *J Food Eng* **2006**, 77, 1012–1017, doi:10.1016/j.jfoodeng.2005.08.032.

19. Fanzone, M.; Peña-Neira, A.; Gil, M.; Jofré, V.; Assof, M.; Zamora, F. Impact of Phenolic and Polysaccharidic Composition on Commercial Value of Argentinean Malbec and Cabernet Sauvignon Wines. *Food Res Int* **2012**, 45, 402–414, doi:10.1016/j.foodres.2011.11.010.

20. Lambrecht, K.; Nieuwoudt, H.; du Toit, W.; Aleixandre-Tudo, J.L. Moving towards In-Line Monitoring of Phenolic Extraction during Red Wine Fermentations Using Infra-Red

Spectroscopy Technology. Influence of Sample Preparation and Instrumentation. *J Food Compos Anal* **2022**, 110, 104542, doi:10.1016/j.jfca.2022.104542.

21. Menezes, J.C. Process Analytical Technology in Bioprocess Development and Manufacturing. In Comprehensive Biotechnology, Second Edition; Moo-Young, M., Ed.; Academic Press, **2011**; Vol. 3, pp. 501–509 ISBN 9780080885049.

22. Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacuring, and Quality Assurance; Services, U.S. Department of Health and Human, Food and Drug Administration, **2004**.

23. Cozzolino, D. Advantages, Opportunities, and Challenges of Vibrational Spectroscopy as Tool to Monitor Sustainable Food Systems. *Food Anal Methods* **2022**, 15, 1390–1396, doi:10.1007/s12161-021-02207-w.

24. Ávila, T.C.; Poppi, R.J.; Lunardi, I.; Tizei, P.A.G.; Pereira, G.A.G. Raman Spectroscopy and Chemometrics for On-Line Control of Glucose Fermentation by *Saccharomyces cerevisiae*. *Biotechnol Prog* **2012**, 28, 1598–1604, doi:10.1002/btpr.1615.

25. Arroyo-Cerezo, A.; Jimenez-Carvelo, A.M.; González-Casado, A.; Koidis, A.; Cuadros-Rodríguez, L. Deep (Offset) Non-Invasive Raman Spectroscopy for the Evaluation of Food and Beverages-A Review. *LWT – Food Sci Technol* **2021**, 149, 111822, doi:10.1016/j.lwt.2021.111822.

26. Teixeira dos Santos, C.A.; Páscoa, R.N.M.J.; Porto, P.A.L.S.; Cerdeira, A.L.; González-Sáiz, J.M.; Pizarro, C.; Lopes, J.A. Raman Spectroscopy for Wine Analyses: A Comparison with near and Mid Infrared Spectroscopy. *Talanta* **2018**, 186, doi:10.1016/j.talanta.2018.04.075.

27. Du Toit, W.J.; Pretorius, I.S. The Occurrence, Control and Esoteric Effect of Acetic Acid Bacteria in Winemaking. *Ann Microbiol* **2002**, 52, 155–179.

28. Guillamón, J.M.; Mas, A. Acetic Acid Bacteria. In *Biology of Microorganisms on Grapes, in Must and in Wine*; König, H., Unden, G., Fröhlich, J., Eds.; Springer Berlin Heidelberg, *2009*; pp. 31–46 ISBN 9783540854623.

29. 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* **2018**, 1058, 9–17, doi:10.1016/j.aca.2018.10.055.

30. Hernández, G.; León, R.; Urtubia, A. Detection of Abnormal Processes of Wine Fermentation by Support Vector Machines. *Cluster Comput* **2016**, 19, 1219–1225, doi:10.1007/s10586-016-0594-5.

31. Bersimis, S.; Psarakis, S.; Panaretos, J. Multivariate Statistical Process Control Charts: An Overview. *Qual Reliab Eng Int* **2007**, 23, 517–543, doi:10.1002/qre.829.

32. Bersimis, S.; Panaretos, J.; Psarakis, S. Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry. *Qual Reliab Eng Int* **2009**, 23, 517–543, doi: 10.48550/arXiv.0901.2880.

33. Muncan, J.; Tei, K.; Tsenkova, R. Real-Time Monitoring of Yogurt Fermentation Process by Aquaphotomics near-Infrared Spectroscopy. *Sensors (Switzerland)* **2021**, 21, 1–18, doi:10.3390/s21010177.

Chapter 5
Conclusions

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

In this chapter, the main conclusions derived from the methodologies developed and the research results obtained throughout the thesis are summarized. The specific conclusions of each individual study have been addressed at the conclusion of their respective research papers.

The main objective of this doctoral thesis was to investigate the winemaking process using vibrational spectroscopy combined with chemometric data analysis. To gain a comprehensive understanding of grape maturity, alcoholic fermentation and maceration processes, multiple sources of variability were examined. Through the definition of experimental designs, it was possible to systematically analyze the impact of various factors on the process and the final product obtained. This approach facilitated the quantification of the individual contributions each factor and provided valuable information on their effects on the winemaking process.

Based on the objectives, specific conclusions associated with the different sections of results are derived:

To assess the impact of several positional factors of grapes in the vine within the same field in grape maturity evolution and grape characteristics.

- The combination of a portable ATR-MIR with appropriate ASCA models can detect the influence of grape position on the bunch and vine on individual grape ripening.
- The evolution of these positional factors throughout the ripening process is highly intricate.

To evaluate how different stages of grape ripeness and winemaking practices affect the chromatic characteristics of final wines using spectroscopic data collected from grapes and the alcoholic fermentation and maceration processes.

- The effect of different viticultural and winemaking practices on obtaining wines with significantly different chromatic characteristics can be evaluated by using MIR and NIR techniques.
- When using ASCA, the spectra obtained from these techniques allow to decompose and quantify the sources of oenological practices variability.

- Grape cultivar and maturity state have a significant influence not only on the chromatic characteristics of the resulting wine but also on the evolution of the extraction of polyphenolic compounds.
- The winemaking practices affect the final product characteristics, but these do not take over the inherent characteristics of the grapes.

To determine the performance of Spatially Offset Raman Spectroscopy to monitor through a container wine alcoholic fermentation and predict main oenological parameters.

- Spatially offset Raman spectroscopy can monitor alcoholic fermentation and predict key oenological parameters associated to the process under real conditions with suspended yeast in the medium.
- The on–line approach (measuring through the container) outperforms the results generated using the at-line approach (taking out an aliquot of the fermentation), as it yields more representative spectra and have less interference coming from fluorescence and scattering.

To develop chemometric strategies to assess the impact of microbiological contamination on the alcoholic fermentation by acetic acid bacteria.

- The alcoholic fermentation signal in the MIR spectra hides the signal of the AAB contamination.
- To enhance the AAB signal in the MIR spectra is necessary to apply diverse unfolding strategies, pre-processing methods, and targeted selection of process stages and spectral regions.
- The application of these techniques makes AAB contamination the main source of spectral information, which allows it to be used to monitor this possible contamination process.

To propose a multivariate statistical tool that enables monitoring wine alcoholic fermentation, detecting sluggish fermentation processes caused by various factors (nitrogen deficit and unwanted temperature changes), and identifying the deviation responsible for the changes in the progression of the process.

- The Evolving Window Dissimilarity Index (EWDI) can be used as an MSPC tool.
- This methodology enables to establish statistical control limits based on fermentations conducted under Normal Operating Conditions.
- Three types of fermentation deviations can be effectively detected: sudden temperature changes, initial nitrogen deficit, and a combination of both.
- The evolution of EWDI over time can detect nitrogen deficit as the primary source of deviation in the fermentation process.

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Appendix

UNIVERSITAT ROVIRA I VIRGILI
QUANTIFYING VARIABILITY IN GRAPE AND WINE QUALITY: A MULTIVARIATE ANALYSIS PERSPECTIVE
Daniel Schorn García

Publications derived from this PhD thesis

[1] D. Schorn-García, J. Cavaglia, B. Giussani, O. Busto, L. Aceña, M. Mestres, R. Boqué. *Microchemical Journal* 166, 106215, (2021).

[2] D. Schorn-García, B. Giussani, M.J. García-Casas, D. Rico, A.B. Martin-Diana, L. Aceña, O. Busto, R. Boqué, M. Mestres. *Foods* 12(5), 962, (2023).

[3] D. Schorn-García, K. Laambrect, B. Giussani, R. Boqué, J.L. Aleixandre-Tudó, M. Mestres. *Manuscript under preparation.*

[4] D. Schorn-García, J. Ezenarro, L. Aceña, O. Busto, R. Boqué, B. Giussani, M. Mestres. *Fermentation* 9(2), 115 (2023).

[6] D. Schorn-García, B. Giussani, O. Busto, L. Aceña, M. Mestres, R. Boqué. *Journal of Chemometrics*, e3465 (2023).

[4] D. Schorn-García, J. Ezenarro, O. Busto, L. Aceña, R. Boqué, M. Mestres, B. Giussani. *Submitted to Food and Biotechnology Process*.

Contributions to national and international meetings attended

Oral Communications

Schorn-García, D.; Mestres, M.; Aceña, L.; Giussani, B.; Boqué, R.; Busto, O.
Desarrollo de modelos multivariantes para el control de la producción vitivínicola. **I Jornadas Científicas de GIENOL**. Online (2021).

Schorn-García, D.; Mestres, M.; Aceña, L.; Giussani, B.; Boqué, R.; Busto, O.
Monitorización in-line de la fermentación alcohólica mediante un espectrofotómetro raman portátil. **XV Congreso Nacional de Investigación Enológica (GIENOL 2022)**. Murcia, Spain (2022).

Schorn-García, D.; Giussani, B.; Busto, O.; Aceña, L.; Boqué, R.; Mestres, M.
Study of grape-ripening process variability using mid-infrared spectroscopy. **In Vino Analytica Scientia (IVAS)**. Neustadt, Germany (2022).

Schorn-García, D.; Ezenarro, J.; Busto, O.; Aceña, L.; Boqué, R.; Mestres, M.; Giussani, B.
A new approach to the Dissimilarity Index as a process control tool to detect deviations in alcoholic fermentation. **XI Colloquium Chemometricum Mediterraneum**. Padua, Italy (2023).

Poster Communications

Schorn-García, D.; Giussani, B.; Boqué, R.; Busto, O.; Aceña, L.; Mestres, M.
Development of multivariate models to monitor and control grape quality and wine production. **I Congreso Anual de Estudiantes de Doctorado**. Online (2021).

Ezenarro, J.; García-Pizarro, Á.; Schorn-García, D.; Mestres, M.; Aceña, L.; Busto, O.; Boqué, R.
Olive ripening assessment methodologies using digital image analysis. **XVIII Chemometrics in Analytical Chemistry (CAC)**. Roma, Italy (2022).

Schorn-García, D.; Ezenarro, J.; Mestres, M.; Aceña, L.; Busto, O.; Giussani, B.; Boqué, R.
Acetic or lactic bacteria contamination? ASCA has the answer. **XVIII Chemometrics in Analytical Chemistry (CAC)**. Roma, Italy (2022).

Ezenarro, J.; Schorn-García, D.; Aceña, L.; Mestres, M.; Busto, O.; Boqué, R. J-Score: a new joint parameter for PLSR model performance evaluation of spectroscopic data. **XI Colloquium Chemometricum Mediterraneum**. Padua, Italy (2023).

Ezenarro, J.; Schorn-García, D.; Aceña, L.; Mestres, M.; Busto, O.; Boqué, R. Studying the nectarine ripening process with near-infrared spectroscopy. **XI Colloquium Chemometricum Mediterraneum**. Padua, Italy (2023).

Ezenarro, J.; Schorn-García, D.; Aceña, L.; Mestres, M.; Busto, O.; Boqué, R. expertPLS: a MATLAB toolbox for spectral preprocessing selection. **XI Colloquium Chemometricum Mediterraneum**. Padua, Italy (2023)

Scientific dissemination activities

Stand in the European Researchers' Night Workshop –Nit Europea de la Recerca

2020. *Descubriendo nuestros sentidos*

2021. *Jugando con las propiedades físicas de los líquidos*

2022. *La Ciencia de los Alimentos*

European Researchers' Night Conference – Conferencia Nit Europea de la Recerca

2021. *Señoría la química es inocente*

Institut Joan Amigó i Callau, L'Espluga de Francolí.

2022. *Señoría la química es inocente*

Institut Josep Tapiró. Reus.

Scientific conferences for schools – Conferencias científicas

Del viñedo a la copa. El científico viaje de una uva

2020. Col·legi Aura, Reus.

2020. IES Blas Cabrera Felipe, Lanzarote.

2022. Institut Tarragona, Tarragona.

2022. Col·legi Sagrat Cor, El Vendrell.

2022. Institut Terra Alta. Gandesa.

Cocina molecular y química

2022. Col·legi Sant Pau Apòstol, Tarragona.

2022. INS Candelera. L'Ametlla del Mar.

2022. Institut Terra Alta. Gandesa.

¿Cocina o laboratorio?

2020. Col·legi Lestonnac – L'Ensenyança

2021. IES Montsià

2022. IES de Flix

Scientific conference Science week – Conferencia científica Setmana de la Ciència

2022. *Cocina molecular y química*

Casal de la Pobla de Mafumet

Centre de Lectura de Reus

Centre Cultural de Cambrils

Advanced course for secondary school teachers – Aprofundiments

2020 & 2021. *La ciencia en una copa de vino.*

Advanced course for secondary school students – Summerlab

2020. *Jugando con las propiedades físicas de los líquidos*

BioNorth Conference – Invited speaker

2021. *Charlas enológicas: Investigación enológica por investigadores de la URV*

Participación “Vols saber que investigo?” – Participation in the "Do you want to know what I am researching?" contest. 2021.