Received: 20 July 2022 | Revised: 7 December 2022 | Accepted: 13 December 2022

DOI: 10.1002/cem.3465

SPECIAL ISSUE - RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Methodologies based on ASCA to elucidate the influence of a subprocess: Vinification as a case of study

Daniel Schorn-García[1] | **Barbara Giussani**[2] | **Olga Busto**[1] | **Laura Aceña**[1] | **Montserrat Mestres**[1] | **Ricard Boqué**[3]

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

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

[3]Universitat Rovira i Virgili, Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Tarragona, Spain

Correspondence
Ricard Boqué, Universitat Rovira i Virgili, Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo s/n, Tarragona 43007, Spain.
Email: ricard.boque@urv.cat

Funding information
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.).

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 preprocessing 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

acetic acid bacteria, ANOVA simultaneous component analysis, portable FT-IR, process deviation

1 | 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 pretreatment. 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, because 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 pretreatment 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 allow 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 preprocessing techniques[8] (i.e., first or second derivatives, Savitzky–Golay smoothing, and multiplicative scatter correction) or combinations of them are usually key steps in the multivariate modeling. 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 situation, 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, and 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 ($p < 0.01$) in some of the i-ASCA intervals, unlike classical ASCA results, which did not show the significance of these factors.[12]

The aim of this paper is to evaluate different methodologies to build ASCA models to study the variation associated with 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 preprocessings to enhance the signals of minor compounds. Finally, different matrix unfolding procedures allowed us to study the individual factors of interest.

2 | MATERIAL AND METHODS

2.1 | 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 a 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 AAB contamination. Spoilage was simulated by adding *Acetobacter pasteurianus* grown in a 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·mL^{-1}.

The fermentation process was kept under a constant temperature of 18°C 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·L^{-1}).

2.2 | 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 Fourier transform infrared (FT-IR) 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 (AAB spoilage)—were analyzed at eight sampling times. Spectra were arranged in a data matrix with dimensions 10 × 845 × 8 (I × J × K: samples × wavelengths × time points).

2.3 | ANOVA-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 ($\boldsymbol{X}_c$) according to (Figure 1)

$$\boldsymbol{X}_c = \boldsymbol{X} - \mathbf{1}\boldsymbol{m}^T = \boldsymbol{X}_{\text{Fermentation}} + \boldsymbol{X}_{\text{Contamination}} + \boldsymbol{X}_{\text{Interaction}} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{X}$ is the original data matrix, $\mathbf{1}$ is a vector of ones, $\boldsymbol{m}^T$ is the mean of all the observations, $\boldsymbol{X}_{\text{Fermentation}}$ and $\boldsymbol{X}_{\text{Contamination}}$ are the matrices representing the effects of each one of the experimental factors, $\boldsymbol{X}_{\text{Interaction}}$ contains the interaction between the factors, and $\boldsymbol{E}$ 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 (five batches per eight sampling times), 40 observations will contain the average profile of NOC

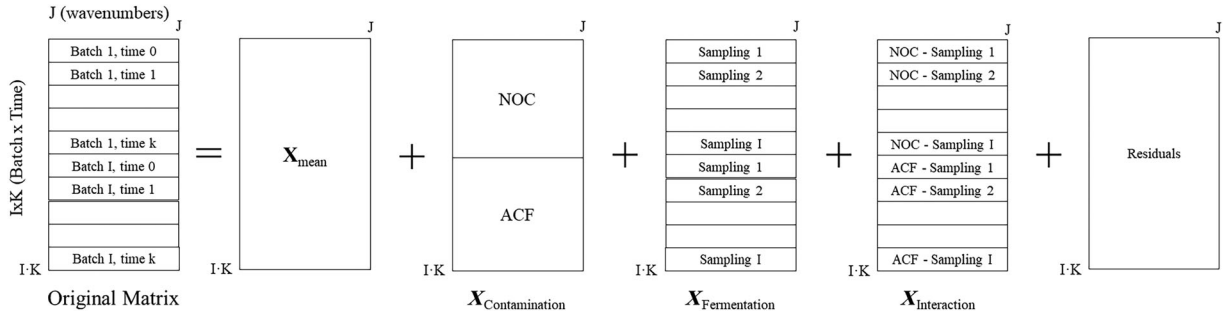

FIGURE 1 Scheme of the procedure applied to build IK × J ASCA models

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. Because 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,

$$\mathbf{SSQ}_{\text{factor } i} = \|\boldsymbol{X}_{\text{factor } i}\|^2, \tag{2}$$

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 with 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 $\boldsymbol{X}_{\text{factor } i}$, the p-value is calculated as

$$p-\text{value}(\boldsymbol{X}_i) = \frac{\text{nbr}\left(\mathbf{SSQ}\left(\boldsymbol{X}_{i,\text{perm}} \geq \mathbf{SSQ}(\boldsymbol{X}_{\text{factor } i})\right)\right)+1}{k+1}, \tag{3}$$

where "nbr" is the number of occurrences, k is the number of permutations, and $\boldsymbol{X}_{i,\text{perm}}$ 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 with 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

$$\boldsymbol{X}_{\text{factor } i} = \boldsymbol{T}_{\text{factor } i} \cdot \boldsymbol{P}^T_{\text{factor } i} + \boldsymbol{E}_{\text{factor } i}, \tag{4}$$

where $\boldsymbol{T}_{\text{factor } i}$ is the score matrix, $\boldsymbol{P}^T_{\text{factor } i}$ is the loading matrix, and $\boldsymbol{E}_{\text{factor } i}$ 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 $\boldsymbol{X}_{\text{factor } i}$ 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., Eagle Rock, 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 10,000,[20] and results were considered to be statistically significant when $p < 0.05$. Throughout the article, only the significant factors will be discussed.

3 | RESULTS AND DISCUSSION

3.1 | 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 h to finish, but to better study the possible contaminations after this process, an additional sampling point (258 h) 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 alcoholic fermentation into acetic acid.[21] However, it has to be pointed out that NOC fermentations also produce low amounts of acetic acid, because yeasts can synthesize this compound to obtain energy as part of their metabolism.[22] Thus, Figure 2B shows that, from 76 h onwards, a different behavior is observed in acetic acid production between NOC and AC fermentations, with a higher production of acetic acid for the contaminated process. At 258 h, a final acetic acid concentration of 1.74 $g \cdot L^{-1}$ was reached, six times higher than for NOC.

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

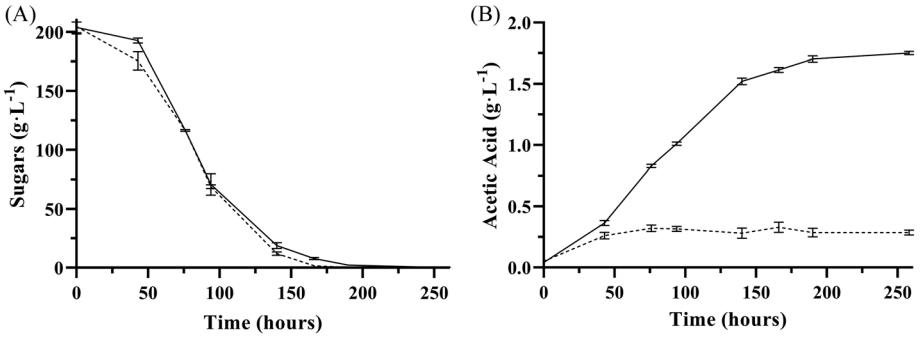

FIGURE 2 Evolution of (A) sugars and (B) acetic acid concentration during alcoholic fermentation. NOC fermentation (dashed line) and AC fermentation (continuous line)

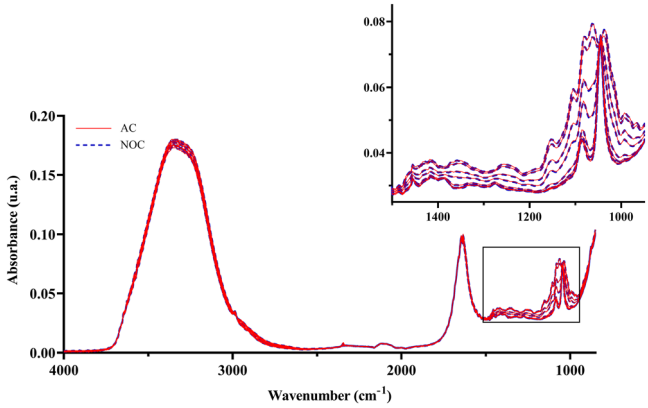

FIGURE 3 Evolution of raw ATR-MIR spectra during alcoholic fermentation and acetification. Upper spectra are a closer overlook of the most variable spectroscopic region. NOC fermentation (blue dashed line) and AC fermentation (red continuous line)

950 and 1500 cm^{-1}. The absorption bands found in this region are related to CH_2, C–C–H, and H–C–O bonds and C–C and C–O stretching vibrations. This region is considered the wine fingerprint region, because 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]

3.2 | ASCA with IK × J 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 (IK × J), to have 80 spectra, which correspond to 10 batches (five NOC and five AC) for eight 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 preprocessing strategies were tested to envision the effect on the factors, and similar results were obtained. The model built with raw spectra (Table 1) will be discussed as example.

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 is 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

TABLE 1 ASCA results for the 80 × 845 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test

	Raw	
Factors	**% Effect**	***p*-value**
Contamination	0.16	0.0001
Fermentation	98.95	0.0001
Contamination × Fermentation	0.11	0.1768
Residual	0.77	

Note: A *p*-value of <0.05 means that the factor is significant.

subprocess, which occurs simultaneously with alcoholic fermentation and produces 1.74 g·L^{-1} of acetic acid. Every data preprocessing tested showed a similar value for each individual effect (Table S1).

Figure 4A shows the score values of the "Contamination" submodel for each sample, colored 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, whereas AC have positive ones. The first loading of the contamination factor (Figure 4B) shows that we can attribute this behavior 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 other 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]

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 behavior for the first component in a PCA analysis of MIR spectra.[9,10] 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 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 preprocessing methodologies, the 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 1795 and 1161 cm^{-1} 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, an IK × J 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.

Despite the fermentation factor is still the main variability source, the contamination factor has gained importance compared with the 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 focusing on a specific region of acids, the alcoholic fermentation remains the main factor.[9] This is because sugars also have major bands in the same region as acids.[24]

3.3 | Time interval ASCA

To deepen 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, an IK × J 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.

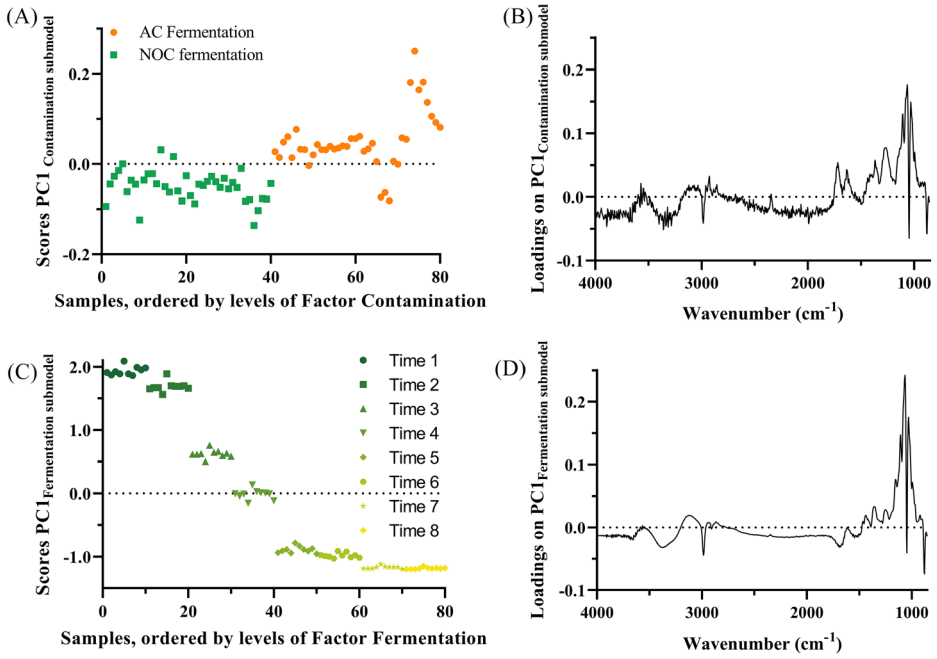

FIGURE 4 ASCA results on the IK × J unfolded matrix (80 × 845). Scores of the first simultaneous component of the (A) contamination factor submodel and (C) fermentation factor submodel, where the samples are ordered by the levels of each factor. Loadings of the first simultaneous component of (B) contamination factor and (D) fermentation factor

TABLE 2 ASCA results for the reduced 80 × 171 matrix, showing the percentage of variance (% Effect) for each factor and the *p*-value resulting of the permutation test

	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 × Fermentation	0.16	0.0003	0.17	0.0003
Residual	0.52		0.68	

Note: A *p*-value of <0.05 means that the factor is significant.

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 % Effect (85.29%) 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 with 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.

Regarding the "Contamination" factor, it is significant from the tumultuous fermentation until the end of the process because acetic acid production increases from 76 h (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 AAB may metabolize sugars at the beginning of the process, because 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

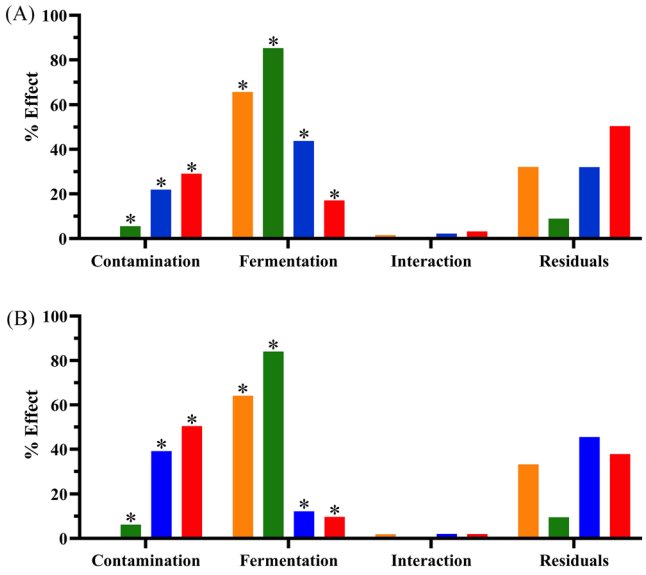

FIGURE 5 ASCA results of the different parts of the process. (A) Four matrices with dimensions 20 × 845 and (B) four matrices with dimensions 20 × 171. Orange: stationary phase, green: tumultuous fermentation, blue: tumultuous fermentation end, red: process end. *: Significant factor

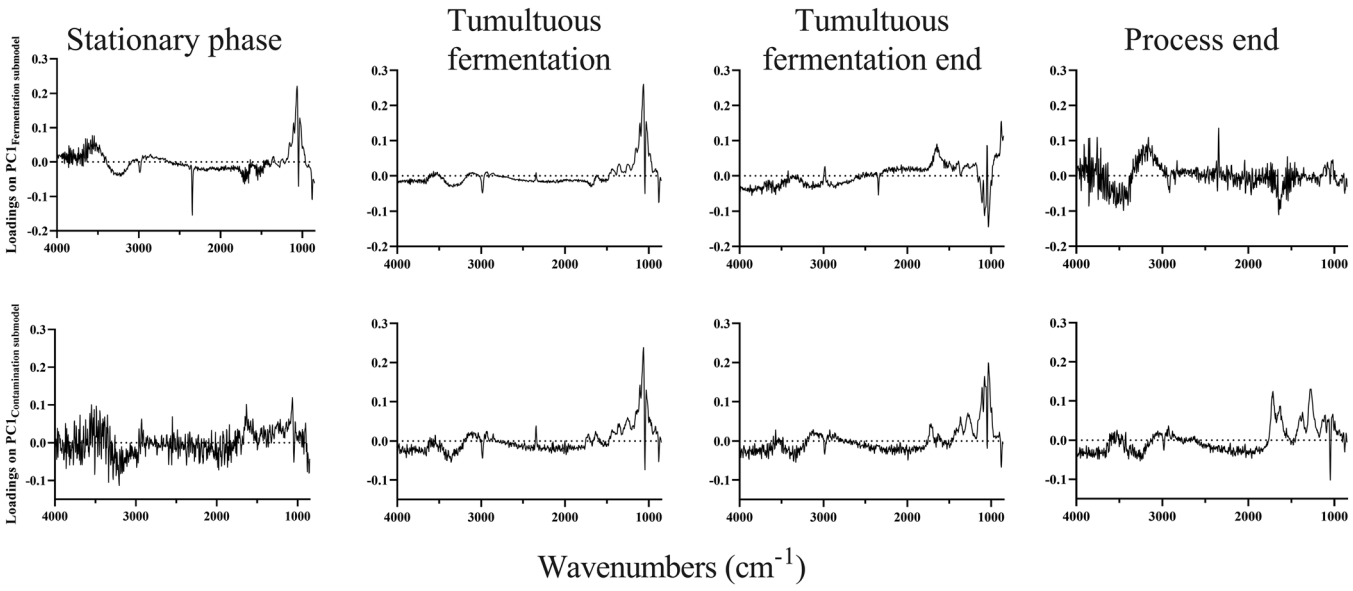

FIGURE 6 Loading of SC1 for fermentation factor (upper row) and contamination factor (lower row) submodels of the ASCA model. Each column represents a phase of the process.

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 behavior 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, because it reaches the highest acetic acid concentration, and the fermentation process has no effect because it has finished.

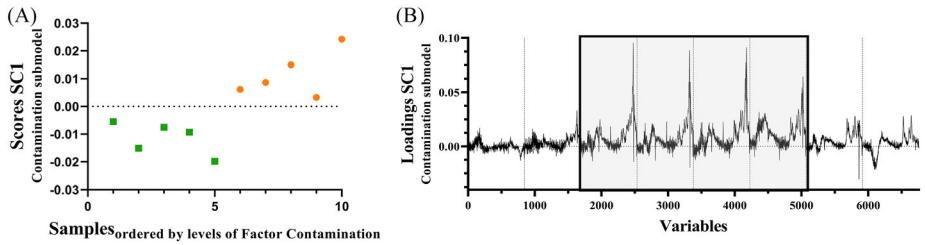

FIGURE 7 ASCA results on the I × JK unfolded matrix (10 × 6760) without preprocessing. Scores of the (A) contamination factor and (B) loading of the first simultaneous component of the contamination factor. Vertical dotted lines divide each sampling point. Areas with larger peaks are highlighted in gray.

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 preprocessing 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 the ASCA model is built using Savitzky–Golay smoothing or the first derivative in combination to SNV preprocessed 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]

3.4 | ASCA with I × JK matrix unfolding

To study the variability of the AAB spoilage, an ASCA model was built on an I × JK 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 eight sampling times (6760 variables) in the columns. In this model, only the "Contamination" factor was considered. Different preprocessing strategies were tested to study their effect on the factor. The ASCA model built using SNV preprocessed spectra (Figure 7) will be discussed as the model with minimum preprocessing. SNV was applied prior to matrix unfolding.

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 preprocessed spectra, respectively. Figure 7A shows the score values for the contamination factor. Different coloring 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 in Sections 3.2 and 3.3, both sugars and fingerprint regions are important parts of the spectra for the contamination factor.

Other preprocessing methodologies were tested in the I × JK unfolding ASCA model, and similar results were obtained. Sight increases in the % Effect for contamination were achieved using smoothing and the first derivative (Table S3) as this preprocessing amplifies smaller peaks, such as the expected for acetic acid in the considered concentration.

4 | 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 AAB 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 into parts and when various spectral preprocessing approaches are applied if required.

ACKNOWLEDGMENTS

The authors would like to thank the Wine Biotechnology group for providing the acetic acid bacteria used in this experiment.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial or personal relationships that could have appeared to influence the work reported in this paper.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3465.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID


Daniel Schorn-García https://orcid.org/0000-0003-0997-2191

REFERENCES

1. Cozzolino D. Advantages, opportunities, and challenges of vibrational spectroscopy as tool to monitor sustainable food systems. *Food Anal Methods*. 2022;15(5):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(3):303-313. doi:10.1016/j.jfoodeng.2007.12.022
3. Bisson LF. Stuck and sluggish fermentations. *Am J Enol Vitic*. 1999;50(1):107-119. doi:10.5344/ajev.1999.50.1.107
4. Guidance for industry PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance. *FDA Official Document*. 2004;(September):2-16.
5. Urtubia A, Pérez-correa JR, Pizarro F, Agosin E. Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. *Food Control*. 2008;19(1):382-388. doi:10.1016/j.foodcont.2007.04.017
6. Chuen Lee L, Liong C-Y, Aziz JA. A contemporary review on data preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemom Intel Lab Syst*. 2017;163:64-75. doi:10.1016/j.chemolab.2017.02.008
7. Miller CE. Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries: second edition. *Chemometrics in Process Analytical Technology (PAT)*. April 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(September 2019):106947. doi:10.1016/j.foodcont.2019.106947
10. Schorn-García D, Cavaglia J, Giussani B, et al. 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 JM, del Olmo A, Engelsen MM, Lundkvist H, Engelsen SB. 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 CB, Rasmussen MA, Amigo JM. Interval ANOVA simultaneous component analysis (i-ASCA) applied to spectroscopic data to study the effect of fundamental fermentation variables in beer fermentation metabolites. *Chemom Intel Lab Syst*. 2017;163:86-93. doi:10.1016/J.CHEMOLAB.2017.02.010
13. Cavaglia J, Schorn-García D, Giussani B, et al. Monitoring wine fermentation deviations using an ATR-MIR spectrometer and MSPC charts. *Chemom Intel Lab Syst*. 2020;201:104011. doi:10.1016/j.chemolab.2020.104011
14. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers RJAN, van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005;21(13):3043-3048. doi:10.1093/bioinformatics/bti476
15. Jansen JJ, Hoefsloot HCJ, Van Der Greef J, Timmerman ME, Westerhuis JA, Smilde AK. ASCA: analysis of multivariate data obtained from an experimental design. *J Chemometr*. 2005;19(9):469-481. doi:10.1002/cem.952

16. De Luca S, De Filippis M, Bucci R, Magrì AD, Magrì AL, Marini F. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD. *NIR and Chemometrics Microchemical Journal*. 2016;129:348-361. doi:10.1016/j.microc.2016.07.021
17. Ryckewaert M, Gorretta N, Henriot F, Marini F, Roger JM. Reduction of repeatability error for analysis of variance—simultaneous component analysis (REP-ASCA): application to NIR spectroscopy on coffee sample. *Anal Chim Acta*. 2020;1101:23-31. doi:10.1016/j.aca.2019.12.024
18. Anderson MJ, Ter Braak CJF. Permutation tests for multi-factorial analysis of variance. *J Stat Comput Simul*. 2003;73(2):85-113. doi:10.1080/00949650215733
19. Zwanenburg G, Hoefsloot HCJ, Westerhuis JA, Jansen JJ, Smilde AK. ANOVA-principal component analysis and ANOVA-simultaneous component analysis: A comparison. *J Chemom*. 2011;25(10):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 MJ, del Carmen G-PM, Troncoso AM, Chen YM, Suh SJ. Acetic acid bacteria and the production and quality of wine vinegar. 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. 2006;1:1-497. doi:10.1002/0470010363
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(1):181-186. doi:10.1016/j.foodres.2010.10.043
24. Bureau S, Cozzolino D, Clark CJ. Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: a review. *Postharvest Biol Technol*. 2019;148(May 2018):1-14. doi:10.1016/j.postharvbio.2018.10.003
25. Moura JCMS, Bonine CAV, de Oliveira Fernandes Viana J, Dornelas MC, Mazzafera P. Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J Integr Plant Biol*. 2010;52(4):360-376. doi:10.1111/J.1744-7909.2010.00892.X

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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. *Journal of Chemometrics*. 2023;e3465. doi:10.1002/cem.3465