Analytica Chimica Acta 1281 (2023) 341902

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Process understanding and monitoring: A glimpse into data strategies for miniaturized NIR spectrometers

Giulia Gorla [a], Alberto Ferrer [b], Barbara Giussani [a,*]

[a] Science and High Technology Department, Università degli Studi dell'Insubria, 22100, Como, Italy
[b] Multivariate Statistical Engineering Group, Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, 46022, València, Spain

HIGHLIGHTS

- Combination of statistical tools for batch processes.
- Insights and drawbacks of statistical tools in NIR process data.
- Miniaturized-NIR spectroscopy data for fermentation.

ARTICLE INFO

Keywords:
Miniaturized NIR spectrometer
Fermentation process
Principal component analysis
Moving-window
Process understanding
Monitoring

ABSTRACT

Background: The implementation of process analytical technologies (PAT) has gained attention since 2004 when its formal introduction through the U.S. Food and Drug Administration was introduced. Manufacturers that need to evaluate the employment of new monitoring systems could face different challenges: identification of suitable sensors, verification of data meaning, evaluation of several statistical strategies to obtain insights about data and achieve process understanding and finally, the actual possibilities for monitoring. Kefir fermentations were chosen as an example because of the chemical and physical transformations that occurred during the process, which could be common to several other fermentation processes. In order to pave the way for monitoring establish the information contained in the data and find the right tools for extracting them is of extreme importance. Strategies to identify different experimental conditions in the spectra acquired with a miniaturized NIR (1350-2550 nm) during process occurrence were addressed.
Results: The study aims to offer insights into good practices and steps to pave the way for process monitoring with handheld NIR data. The main aspects of interest for batch processes in preliminary evaluations were investigated and discussed. On the one hand, process understanding and, on the other, the possibilities for process monitoring and endpoint determination were examined. The combination of different statistical tools allowed the extraction of information from the data and the identification of the link between them and the chemical and physical changes during the process. In addition, insights into the spectra characteristics in the studied spectroscopic range for kefir fermentation were reported.
Significance: The capabilities for miniaturized NIR spectra to represent and statistical strategies to characterize different experimental conditions in a real case fermentation occurrence were proved. The strengths and limitations of some of the common approaches to catch changes in fermentation condition were highlighted. For the various statistical approaches, the chances offered in the research and development stages and to set the scene for monitoring and end-point detection were explored.

1. Introduction

To date, process analytical technologies (PAT) are used for real-time process or product monitoring in several industries, as well as for raw material identifications and quality control. The main goals of PAT are product quality assessment and process understanding which could lead to higher process efficiency [1]. In this context, using molecular spectroscopies such as near-infrared spectroscopy is one of the tipping points

* Corresponding author.
E-mail address: barbara.giussani@uninsubria.it (B. Giussani).

https://doi.org/10.1016/j.aca.2023.341902
Received 29 April 2023; Received in revised form 28 September 2023; Accepted 10 October 2023
Available online 10 October 2023

to pave the way for industry 4.0 application. The acquisition of real-time process information can be achieved by implementing PAT during the manufacturing process. Currently, some NIR technology has been implemented in food, petrochemical, and pharmaceutical industries mainly because of the promising results obtained with fiber optics [2–6]. In this context, batch processes are ubiquitous and several statistical strategies were considered over the years in the literature [7,8].

Driving causes of the expansion of NIR technology result from the recent trend towards miniaturized and handheld spectrometers based on MEMS and MOEMS [9]. The portability of this instrumentation can be seen as based on two goals: transferability of analysis from the lab to the field and on-line applications [10]. Several studies have largely discussed noticeable results at industrial and lab scales with high interest [11,12]. Conversely, miniaturization carries restricted spectral ranges and lower performance than its laboratory counterpart [13]. A field where the advantages of miniaturized spectroscopic methods are manifold is bioprocess monitoring. Indeed, when coupled with chemometric tools, aside from including real-time capability, they offer a non-destructive nature, easy maintenance, the possibility for simultaneous determination of multiple target analytes in complex fermentation broths and last but not least, affordability [14,15].

Undeniably, NIR spectroscopy has already been proven suitable for dairy fermentation processes [16–21]. However, it is worth noting that previous works were conducted mainly with benchtop instrumentation which is robust and technologically mature. Acquisition configurations suitable for offline, at-line and online analysis with probes were evaluated. Considering the differences in costs, primary work environments, and process implementation possibilities, the previous studies are far from comparable to those that imply miniaturized instrumentations [22]. Together with possible manufacturing improvement, there is still a lack of knowledge about miniaturized NIR applications and how to extract data information efficiently. Only recently, it arose that developing ad hoc strategies to obtain valuable information from a miniaturized NIR spectrometer implies multiple steps: from reliable data acquisition strategies to suitable data analysis and interpretation [22,23].

In this study, a dairy fermentation process was investigated through spectroscopic measurements under different experimental conditions. Spectra acquisitions were conducted in the range of 1350–2550 nm with a MEMS-based instrument by simulating, at a lab scale, a reflection geometry on a glass window of a bioreactor. In a previous study, the identification and evaluation of different set-ups for data acquisition was conducted as first step to pave the way for process understanding and monitoring [24]. After that, in this research, different process conditions were tested and the crucial aspect of assessing the suitability of the methodology for gaining process information was addressed.

The main aspects offered when coupling spectroscopic data and statistical analysis were considered: from process understanding to possibilities for monitoring and endpoint detection. Different conditions of the process occurrence were investigated in order to exploit and evaluate the possibilities offered by different statistical approaches on data from a miniaturized NIR. Nowadays, the typical issues that relate to industries deal with the extraction of information from big quantities of historical data but some strategies for preliminary investigation could still be useful.

The aim of the study is manifold. A methodology is proposed to investigate spectroscopic data to exploit the information from different perspectives when the possibilities of data acquisition are limited: a crucial aspect is the comparison and combination of the results obtained with different approaches. The coupling of data from a miniaturized spectrometer and multivariate strategies was considered to focus on different goals and approaches and to guarantee the extraction of the maximum amount of information. Indeed, different multivariate statistical approaches are applied and then discussed by highlighting the strengths and limitations of the strategies when applied to the real case example. Since one of the essential keys to monitoring a process is its deep understanding and knowledge about data acquired, the main traits related to the spectra registered during processes are explored and commented. Insights about the kefir fermentation process are enhanced. The relation between the process occurrence at different temperatures and the spectra acquired with and without sample pretreatment is investigated. Due to instrumental limitations, obtaining insights related to different experimental conditions with miniaturized spectrometers is not obvious; the sensibility to such changes could be a pivotal aspect to early-detect deviation from the normal operating conditions. In addition, this research aims to demonstrate the potentiality and drawbacks of miniaturized NIR instruments data when coupled with statistical tools. Clues are provided to decide whether a tool can be useful and to what extent when coupled with spectroscopic measurement.

2. Materials and methods

2.1. Batch kefir production

Semi-skimmed fresh milk was chosen as fermentation media as is one of the most common. The main nutritional characteristics are 1.6 g of total fat, 3.4 g of protein and 5.2 g of carbohydrates (Table S1). It was purchased in an Italian supermarket (Como, CO) and stored in the refrigerator at 4 °C until further use. Active kefir grains were obtained from Kefiralia (Burumart Commerce SL) and conserved in fresh whole milk by changing the media each two days for one week before starting the analysis.

For each fermentation, 100 mL of milk was pre-warmed in a laboratory heater and then used as fermentation media. Milk was inoculated with 10 g of kefir grains extracted from the colony conserved in fresh milk and weighted. The fermentation was performed in a beaker by maintaining the pre-heated temperature by a cooled incubator FOC 225I (Velp Scientific). Two different temperatures of fermentation were tested: 25 °C and 30 °C. At both temperatures, a control batch containing only milk was also conducted and observed. pH values were obtained at the inoculation time (t = 0) and at each sampling point (every 15 min over the fermentation) by a pH/ORP Meter HI9125. Data for some processes were acquired by stirring the fermentation media before spectra acquisition (pretreated samples) while, for other processes the fermentation containers were positioned on the spectrometer without pretreatment (samples with no pretreatment). The experimental plan is reported in Fig. 1. Three fermentation replicates (3 x batch) were obtained at each temperature and mixing condition and were conducted over different days. The fermentations were considered to have ended at a pH of about 4.6, but the spectra were acquired until a pH of about 3.2, and at that point, the grains were removed and deposited in fresh milk with the rest of the colony for subsequent use. The times of fermentation processes occurrence were found to not be significantly different depending on the experimental conditions.

2.2. Instrument and materials

Spectra were collected with a NeoSpectra-Scanner (Si-Ware System), which covers the range of 1350–2550 nm (7407 cm^{-1} - 3922 cm^{-1}). Data were autosaved in the default format ".Spectrum" and then were converted to.txt files and processed with MATLAB (9.10.0.1602886 - R2021a). Spectra of 5 experimental replicates (obtained with sample repositioning) were acquired at each sampling time (every 15 min over the fermentation) during each fermentation replicate (each batch). A 100 % reflectance reference was performed with a Spectralon® approximately 15 min after the spectrometer was turned on and before each sampling time. A time scan of 5 s was selected. 250 mL borosilicate glass 3.3 beakers (Rasotherm®) were used as fermentation batch and for spectra acquisition. Control batches without grains inoculation were acquired at the same intervals. Beakers were gently placed on the spectrometer window from the bottom (Supplementary Fig. S1).

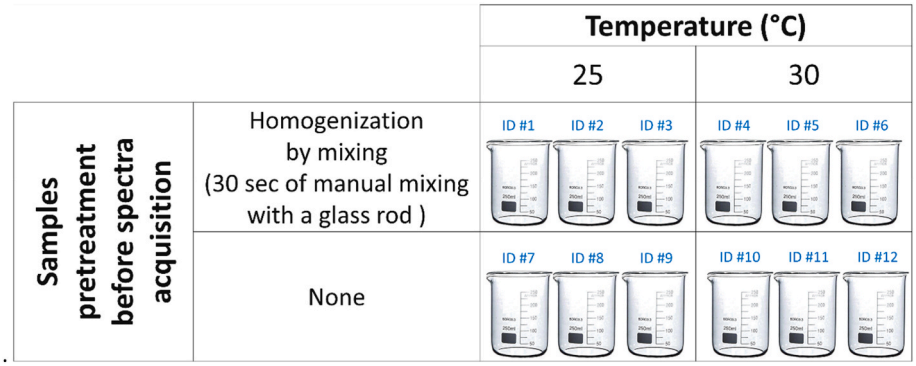

Fig. 1. Experimental plan for fermentation processes.

2.3. Statistical data analysis

Data elaboration and analysis were performed with in-house routines programmed in MATLAB 9.10.0.1602886 - R2021a (Mathworks Inc., Natick, MA, USA) and PLS toolbox 8.9.1 (Eigenvector Inc., Manson, WA, USA).

Each fermentation batch was organized as **X** matrix with time of sampling in the *n* rows and *j* wavelengths in the columns. In addition, as batch processes, the fermentation under investigation, could be intended as a three-way array $\underline{\mathbf{X}}$ containing *j* variables (74 data points wavelengths) measured at each *k* sampling time (34 points) in I batches (12 processes obtained from 2 temperature conditions x 2 type of pretreatment x 3 fermentation replicates). The measures of pH values for replicates were reported in a **Y** matrix (batches x times).

The reproducibility of spectra and the changes in sample heterogeneity through the fermentations were investigated by root mean square error (RMSE) statistics on each batch [25,26]. The RMSE was calculated as:

$$\mathrm{RMSE_r} = \sqrt{\frac{\sum_{j=1}^{n} \mathrm{D_{jr}^2}}{\mathrm{n}}} \text{ with } \mathrm{D_{jr}} = \mathrm{y_{jr}} - \bar{\mathrm{y}}_{\mathrm{j}} \quad (1)$$

Where $\mathrm{y_{jr}}$ is the reflectance at λ_j for experimental replicate *r*, $\bar{\mathrm{y}}_{\mathrm{j}}$ is the reflectance at λ_j for the mean spectrum of R subsamples of a sample, and *n* stands for the number of data points in a spectrum. The index r is used to indicate that the error is calculated within experimental replicates. Indeed, the statistics were calculated considering the reproducibility at each time of fermentation and considering five experimental replicates as subsamples.

Principal Component Analysis [27] on the average spectra of the 5 experimental replicates was carried out on each $\mathbf{X}_i$ to identify the main features contained in the data related to the fermentation occurrence. Data were mean-centered.

Multivariate statistical tools were intended to consider time dependency and dynamics over the process which, could be assumed, happens through different phases. Data analyses were performed maintaining the experimental replicates as individual spectrum in $\mathbf{X}_i$ and then, averaging the experimental replicates. On each $\mathbf{X}_i$, moving block standard deviation (MB-STD), as expressed by Sekuic et al. [28], was considered to identify trends and trajectory of the variability in the spectra. A key parameter to identify changes is the window size (w): sizes from 5 to 25 were tested. A visual explanation of the strategy is reported in Supplementary Material Fig. S2.

Moving window principal component analysis was applied (MWPCA) [29] on each $\mathbf{X}_i$ to elucidate details and spectra interpretation. The statistical process monitoring method was introduced on simulated data by Kano et al., in 2001 [30]. A time-window (of *hs* size) slices the dataset under investigation and subsequently streams in the time direction *hs* varying from 5 to 15 were tested. At each step of the window slicing, a PCA is calculated and the loadings (***p***) are used to determine the so-called dissimilarity index (A_i) between steps. A_i is calculated as the difference between 1 and the inner product of transposed loadings obtained at a defined step (*t*) and the one chosen as reference ***p***(0):

$$A_i = 1 - \left|\boldsymbol{p}(t)^T \boldsymbol{p}(0)\right| \quad (2)$$

In other words, a PCA model is obtained at a defined time-window and so is ***p***(*t*), which represent the current operating condition. Then, the value at the *t* step are compare to its reference ***p***(0), obtained from the first PCA model, the one which include the starting condition of the process. If the values are equivalent, A_i become zero, while if the loadings ***p***(*t*) are orthogonal to ***p***(0), A_i become 1. The dissimilarity index allows to identify trends in the process and changes in process direction and so different phases, intended as different steps during the fermentation when the loadings represent different variation. To better highlight the region of the spectra changing in each phase, loadings variance during different phases of the process were calculated (e.g., the variance calculated over loadings from ***p***(0) up to ***p***(*t*) with *t* identified based on the pH values trends and the dissimilarity index curve). Loadings variance is intended as the squared standard deviation of loadings during each phase. A visual explanation of the strategy is reported in Supplementary Material Fig. S3.

Considering the 3D nature of the data consisting in batches × time × wavelengths, the unfolding of $\underline{\mathbf{X}}$ was considered in different directions. The batch-wise unfolding (BWU) [7] is conducted in order to obtain an the array sliced in the time direction, while the variable-wise or observation-wise unfolding (OWU) [7] is obtained with each batch matrix placed under another (Supplementary Material Fig. S4). PCA [31] was first applied to the unfolded data. Then, the pH measured during the acquisition of spectra was considered as the process's maturity index, and Partial Least Squares models [32] were developed to correlate spectra to pH.

3. Results and discussion

3.1. Spectra

After having set the acquisition strategy, a first approach could consist in verifying that the data acquired by the instrument effectively contained the information needed for process monitoring. Good practice consist in looking at the spectra characteristics. The reflectance NIR spectra of four different batches are shown in Fig. 2. Initially, the fermentation media is liquid milk with a determined concentration of sugar. Ultimately, it is a viscous substance with different hydrogen interactions, acidic conditions, and protein type of aggregation [33]. External reflectance measurements were elected because milk's viscosity increased during the bioprocess's progress, leading to a product with

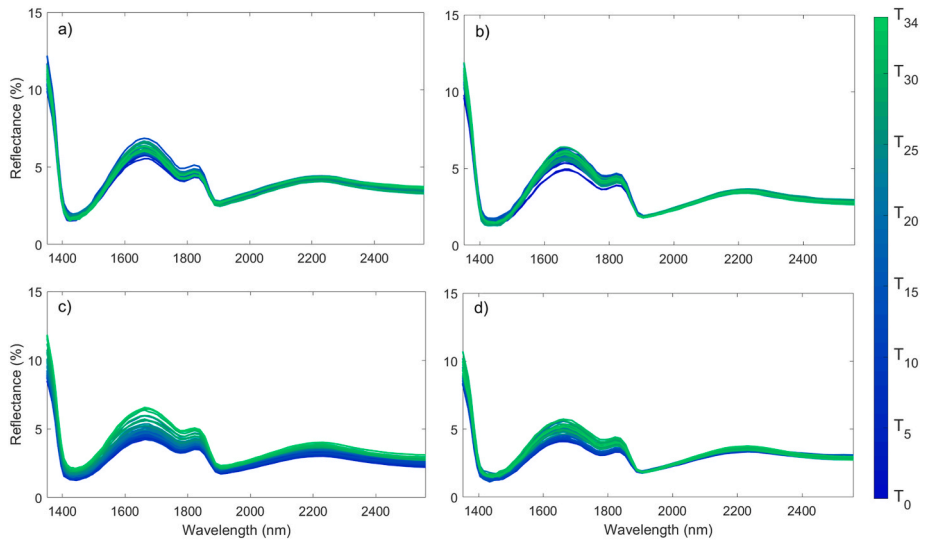

Fig. 2. Example of raw spectra of a kefir fermentation acquired for 8 h and 30 min at interval time of 15 min 35 time points were acquired with T_0 equal to the time of inoculation: a) batches conducted at 25 °C by mixing the samples before spectra acquisition, b) at 30 °C by mixing the sample before spectra acquisition c) at 25 °C by placing directly the beaker on the spectrometer window without stirring the sample d) at 30 °C by placing directly the beaker on the spectrometer window without stirring the sample.

a weak gel behavior [34]. The physical and chemical changes and the sensibility of the techniques set the scene for good perspectives of process monitoring.

As the first approach, spectra appearances were inspected and interpreted. The first overtone of O–H stretching and the combination region of O–H stretching and bending are visible around 1450 nm and 1910 nm. The 1350-1360 nm spectral region can be assigned to proton hydrates/solvated hydronium ions and/or water hydration shells [29] even if the interpretability of this region is limited due to the vicinity with the limit of the spectrometer acquisition range. A band around 1800 nm could be attributed to the combination of the C–H stretching first overtone and C–O second overtone [35]. The regions between 1550 and 1780 nm and around 2200 nm could be ascribed to a combination of O–H and C–H tones and the first overtone of O–H and C–H bonds. Another spectral region of interest, where combinations and the first overtone for O–H bonds appear, is the range between 2030 and 2130 nm and the one around 1500 nm.

Despite the common trends, the spectra interpretation could be slightly diverse for different conditions regarding spectra intensity and peak shifts. In general, a reflectance intensity between 0 and 15 % is reported and the main traits of kefir spectra could be identified and interpreted for each fermentation condition. Across different conditions, the principal variation observable is related to the shift of the baseline, ascribable to the different scattering properties of the fermentation media along with the time of sampling. Indeed, the spectra acquired over a fermentation process after gently mixing seem to be less variable with a percentage standard deviation between 2 % and 7 % for 25 °C and between 1 % and 15 % for 30 °C, whereas a percentage standard deviation between 4 % and 16 % for 25 °C and between 2 % and 22 % for 30 °C is reported for static samples. In addition, fermentation batches conducted either at 25 °C and 30 °C without stirring clearly showed an increasing trend of the baseline over times.

3.2. Enhance process dynamics in the spectra

Once the investigation of spectra has been conducted, several approaches could be used to investigate the evolution over time of the signal related to the fermentations. As emerged from the literature [36], the information in NIR spectra is usually very complex, self-correlated, and spread over the spectrum. In addition, physical properties such as compactness, particle size, particle distribution, and matrix contribution usually affect the spectra. Within this frame, despite the visible trends over the spectra, the univariate approach has already been proven inadequate to investigate the transformation occurring during kefir fermentation. On the other hand, the application of exploratory analysis tools as principal component analysis could allow the identification of trends and main features related to the occurrence of fermentation under defined conditions with several strategies of spectra acquisition [24].

3.2.1. Replicate spectra heterogeneity

While fermentations proceed, complex bio-chemical transformations occur, and the viscosity of the sample change. The starting material for the fermentation, commercial fresh milk, has intrinsic heterogeneity that would change during the process to become a more viscous product. NIR spectra usually are sensitive to changes in the physical properties of the samples as viscosity and heterogeneity [37]. The heterogeneity reflected in NIR spectra of different food samples could be evaluated through RMSE statistics (Equation (1)) [25,26]. The mean and standard deviation of RMSE statistics calculated on 5 experimental replicates over the different acquisition times of the process were investigated (Supplementary Fig. S5). RMSE statistics allow evaluation of the measurements' repeatability and could provide insights about the error intrinsically present in the spectra. The inspection of the results appeared mainly dependent on the batch under analysis. As in this case study, trends that do not follow a specific and reproducible pattern for the RMSE statistic could suggest the joint influences in the spectra reproducibility of physical and chemical changes. As could be expected, the RMSE values of homogenized batches are, on average, characterized by higher values of RMSE mean than the unhomogenized ones as the caught heterogeneity is bigger, as was verified through ANOVA by obtaining a p value < 0.05. This could be ascribed to the different intensities and the different reproducibility of the spectra over time, depending on the conditions. Due to these results, it is possible to state only using this statistical tool that the parameters changed during the experimentation influenced the spectra results. RMSE statistics were also obtained for the control batches by obtaining values comparable to those of samples acquired without previous pretreatment (Supplementary Fig. S6).

This approach could certainly be used as the first screening to understand the main characteristics captured by the spectrometer during the process under study with major influence on the spectra. Based on

simple subtractions, mean, and standard deviation calculations, it is easy and immediate. A clear visualization through bar plots allows exploring the repeatability of the measurements during the process and between processes, easily allowing the comparison of intrinsic spectra reproducibility in different conditions. If considering a first approach with the instrument or data analysis, the tool could result in interesting preliminary insights. Regarding monitoring, the issues of this method relate to the fact that no information about the reason for the reproducibility changes is available. The spectroscopic area of interest or the type of variation in the spectra (in baseline or peaks intensities) are not really at disposal. Concerning the endpoint detection of the process, this strategy resulted limited and other strategies are needed.

3.2.2. Principal component analysis

An interesting tool for feasibility studies is represented by principal component analysis. NIR spectral data reflect the typical multicollinearity of infrared spectra [36,37]. For this reason, in order to investigate the main trends in the spectra of each fermentation and identify important variables and their relations, multivariate exploratory analysis was conducted through principal component analysis on each fermentation batch and the results were investigated from several angles. Since keeping the original baseline shift and the physical information related to scattering resulted as a crucial aspect in a previous study [24], no preprocessing except for mean-centering was applied to the spectra.

The explained variances of the first and second principal components of the models obtained were reported in Table 1. As could be noticed, the first PC explained most of the information in the data. The values of the second PCs are always inferior to 7 %, indicating the presence of some information that could depend on the dataset under consideration (fermentation variability).

After that, the scores for each batch processed at each experimental condition were plotted together. Investigating the score (Fig. 3) and loading plots (Supplementary Fig. S7), it is possible to obtain information about the process occurrence in each fermentation, the spectral region that changes, and the reproducibility of the experiments. It should be pointed out that the scores assumed a different range of values depending on the condition of fermentation and acquisition.

First of all, depending on the pretreatment of the sample before the spectra acquisition, different behavior of the scores can be identified. In the case of sample mixing, a rapid increase with subsequent stabilization of the scores could be noticed. A similar trend was recognized for the lactobacilli and yeast growth during fermentation by Guzel-Seydim et al. [38]. A possible explanation is that when the sample is mixed before the spectra acquisition, the fermentation could be partially accelerated by inoculation of oxygen and the fermentation nucleus could spread over the samples. According to this, the plateau in the scores tendency is obtained when the sample is homogenously fermented. Within this frame, the identification of the end-point, despite the noisy tendencies, is still conducible to the scores behavior. The end-point could be identify when the curves reach a plateau structure.

The spectra acquired directly without pretreatment showed a different trend. A possible reason for these trends could be identified in the changes in the jellification characteristics of the medium. If the fermentation is carried on without mixing, it starts near the grains, and it is observed that a viscous product is formed at the bottom of the beaker after some hours. The scores reflect the experiment observations by testing the effectiveness of the spectrometer to look into these kinds of changes. The spectra catch both physical and chemical information about the process. It could be interesting to note that spectra include not only process behavior but also information about the variability in spectra acquisition, as seen in the case of Fig. 3 c), where an outlier point could be detected at time 20 for batch #7. As the subsequent acquisitions of the same batch followed the typical trend for the experimentation conditions, an acquisition problem for the spectra at that time is assumable.

In literature, the temperature was proven to influence the occurrence of the fermentation, the bacteria and yeast functioning, and grains growth [39], and it greatly influences also NIR spectra [40]. According to this, we could identify higher slope and lower reproducibility in the scores when the fermentation was conducted at 30 °C with respect to their counterpart at 25 °C. The variability across fermentation replicates was lower at 25 °C and without sample pretreatment. This could be due to two main facts: the main spectral differences at initial and final acquisition times that lead the principal component axis in the same direction and the easiness of experimental control at those conditions.

Looking at the loadings, the main spectral changes occurred in common regions across conditions. Despite that, it could be stated that at 1822 nm and around 2200 nm the loadings of the spectra of the unmixed sample are bigger in absolute value than for the homogenized samples resulting in differences at shoulders levels around 1780 nm (ethanol concentration) and 2030 and 2130 nm (O–H bonds). PCA models for control batches were calculated and no specific trends in the scores plot were identified other than a normal data distribution (results not shown).

Overall, using these statistical analyses and visualization tools is powerful in evaluating each process's singular tendency and obtaining a general look at the differences between replicates and conditions. The results of this statistical approach allow for better characterizing the differences in the spectra related to the different conditions of experimentation.

It is worth noticing that principal component analysis aims to identify the direction of maximum variance in the data under investigation, so here the maximum variance contained in each individual experiment is looked into. PCA represents a powerful instrument for feasibility studies, preliminary investigations and process understanding. The procedure aims to exploit information related to each experiment and it is particularly effective when the number of batches to study is limited. On the contrary, in a monitoring phase, a direct comparison of the scores obtained from the same decomposition (same PCA model) with common loadings is recommendable to estimate the effective variation between and within batches. Indeed, the comparisons obtained from distinctive models furnished insights on the experiments but did not allow to establish the typical behavior or the control limits for a control chart.

Of particular interest was also the observation of different replicates obtained under the same conditions. It could be expected that, in fermentation processes the duration (in time) varies from batch to batch [41]. Within this research, the observation of pH values as standard reference show differences in fermentation time depending on the temperatures but not within batches conducted under the same

Table 1
Explained variance of principal components analysis models obtained for each fermentation process.

BATCH ID	PC1 (%)	PC2 (%)	Total variance explained	BATCH ID	PC1 (%)	PC2 (%)	Total variance explained
1	91.71	3.67	95.38	**4**	91.76	3.28	95.04
2	95.94	1.19	97.13	**5**	94.20	1.96	96.16
3	90.39	3.41	93.80	**6**	84.97	6.72	91.69
7	98.14	0.64	98.78	**10**	95.37	2.16	97.53
8	88.92	3.57	92.49	**11**	99.04	0.39	99.43
9	96.02	1.62	97.64	**12**	87.81	4.51	92.32

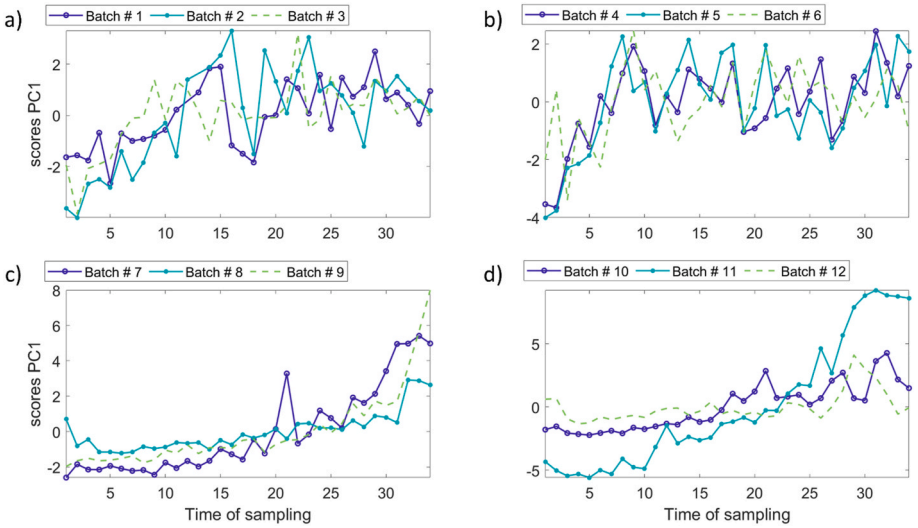

Fig. 3. –Resulting scores on the first principal component for PCA models conducted on each fermentation: a) batches conducted at 25 °C by mixing the samples before spectra acquisition, b) at 30 °C by mixing the sample before spectra acquisition c) at 25 °C by placing directly the beaker on the spectrometer window without stirring the sample d) at 30 °C by placing directly the beaker on the spectrometer window without stirring the sample. Plots are represented within different ranges of y values to allow the focus on the trends in each type of experiment more than the direct comparison of scores values.

conditions. Indeed, apart from the variability of the scores within the same PC (vertical oscillations), that could be conducted to the experimental variance included in the data, the behavior of the scores tendency were substantially aligned.

3.2.3. *Dynamic modelling*

Considering the complexity of the kefir fermentation process that take place in different phases [33,42], an advisable approach could reside in using dynamic modeling that enhances and elucidates the details of physical and chemical changes during the process occurrence instead of investigating the data as whole from the beginning to the end of their acquisition. Three approaches were applied for this kind of modeling. According to the dataset, the algorithms were slightly modified to consider the replicates' presence. Dataset A consisted of all the spectra acquired and disposed as a singular line of the matrix (170 × 74 for each batch). Dataset B was constructed by considering each line of the matrix under investigation as the average spectrum of the five experimental replicates (34x74 each batch). Dataset C consists of the standard deviation calculated over the five experimental replicates at each time (34x74 each batch).

3.2.3.1. *Moving block standard deviation.*

As the spectral repeatability was proved to change over the process due to different contributing factors (physical and chemical changes), moving block standard deviation was studied. The outcome of the calculations on datasets A and B did not show a significant tendency or trajectory related to the fermentation process (results not shown). Different insights were achievable while evaluating the results for dataset C. The window w = 12 was chosen because of the immediacy in interpreting the results obtained. In Fig. 4, the moving block standard deviation is represented vs. the times of the process group for each type of sample pretreatment and temperature of the experiment.

Observing the plots obtained for homogenized samples at 25 °C and 30 °C, the trajectory related to the process occurrence is identifiable. The standard deviation values in the moving window decreased with different variability depending on the temperature. Chemical changes

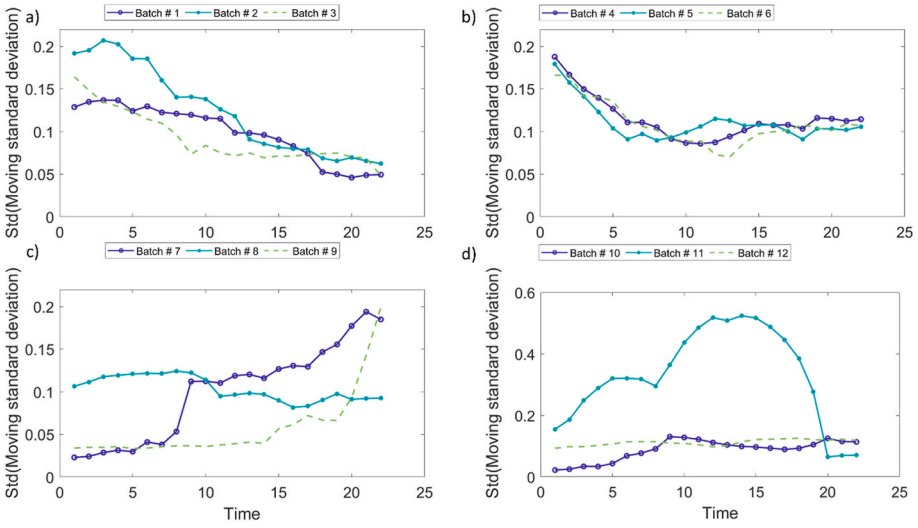

Fig. 4. –Moving block standard deviation over the fermentation process: a) 25 °C and pretreated samples, b) 30 °C and pretreated samples, c) 25 °C and samples without pretreatment, d) 30 °C and samples without pretreatment. The scale on the y axis is different for d) to allow the visualization of the anomalous batch.

and physical heterogeneity of the samples are caught, and the standard deviation over time is representative of the process occurrence. On the contrary, the plots associated with the spectral acquisition without a mixing step do not carry clear information on the process advancement. A likely reason for that could be attributed to the loss of information in the spectra related to physical changes. In other words, if the samples are not mixed, a layer of viscous product deposits on the bottom of the beaker right after the start of the fermentation preventing the spectra from being representative of the physical changes in the whole beaker. The repeatability of the results seems higher when the sample is pretreated, and the temperature of fermentation is 30 °C. In addition, with this statistical tool, batch #8 and batch #11 resulted in anomalous tendencies with respect to the others obtained under the same fermentation and acquisition conditions. The control batches did not show significant tendency and reported values, on average, below the ones of the fermentations (results not shown).

For monitoring purposes, results seem promising only in some experimental conditions. The different behaviors related to different conditions allow to prove the feasibility of fault detection. When the tendencies are identifiable, the standard deviation tends to a *plateau*, which could be considered the point at which no remarkable changes are captured by the spectra. The cross-reference with pH value allows to consider the stasis of the standard deviation along time as the endpoint of the fermentation. However, it must be taken into account that the exact criteria to apply when using MB-STD for endpoint detection must be bent to the production needs, limits, and specifications.

An interesting comparison could be identified between these results and those from singular PCA models. PCA models enhance the process occurrence for pretreated batches better, while MB-STD stressed the process for samples without pretreatment. The intrinsic characteristics of the statistical tools seem to allow the enhancement of the fermentation process based on different data characteristics.

In PCA, the variance between spectra analyzed after being acquired at different times allows for identifying the differences between the starting and ending conditions. The difference between subsequent mean spectra is emphasized.

MB-STD takes into consideration the evolution of the spectra standard deviation. The main characteristic of this statistic is to allow the visualization of step-by-step changing in the spectra with no information on the spectral range interested in the modification. The benefit of this approach is that, depending on the window size, it allows to take into account the past and present conditions at each time step. Indeed, at each calculation, the information carried by the data is incorporated into that present in the previous steps to calculate their relationships in terms of standard deviation. Overall, this statistical approach allowed the identification of other spectral characteristics related to the changes during the fermentation and could furnish information about abnormalities.

When evaluating this method for future monitoring, it is necessary to consider the need for choosing different parameters. Even though the most important parameter is the size of the window width (w) over which the calculation of the standard deviation is conducted, how to intend the matrix over which let the window slide (maintaining or averaging replicates) could also be important. The evaluation of these parameters is strictly related to the number of acquired times, the process duration, and the proper characteristics of the process under investigation. A suitable practice could consist of evaluating the interpretability of the results by trial and error and with the help of a plant expert. It must be noted that, in an industrial environment, a key aspect to consider could be the need and limits for punctual information (acquisition frequency, density, and granularity) and the possibilities for prompt intervention. Other than that, there are no possibilities for chemical or direct physical interpretation, while these could be considered a relevant step for considering corrective actions toward eventual faults. It is impossible to easily know what region of the spectra or effect is responsible for the changes in the moving standard deviation in each step. The possibilities are several: peak shifts, baseline shifts, and different chemical information.

The simple univariate visualization could, under some conditions, allow determining the end-point of the process, and failures in the homogenization system would show the loss of the monitoring information. An interesting output of these results is that the chemical and physical changes occurring during the fermentation are already summarized with only one index. The data dimensionality is reduced, and there is no need for excessive historical data to establish a typical process behavior (Normal Operating Conditions). Indeed, it is possible to identify a limit of the value at which the fermentation could be completed. The need for further acquisition to construct a control chart is limited and could pave the way for immediate end-point identification and faults detection. Other historical data to consider as normal operation condition (NOC) batches could allow a more in-depth study of the process variability.

3.2.3.2. Moving window – PCA. Typically, fermentation and batch bioprocesses occur through different phases according to the microbial growth [43] and dynamic modelling could represent a powerful tool to enhance the understanding of each phase and obtain a more sensitive monitoring system. The microbial growth is typically related to the pH values. As in the case of yogurt [29], three stages could be identified in the kefir fermentation by investigating the pH changes (Supplementary Fig. S8). Since it is not obvious to identify the same changes through the spectra of NIR miniaturized spectrometers and to interpret spectra changes occurring during different phases, moving window principal component analysis (MWPCA) was applied [29]. In this case, datasets A and C furnished confused results with no clear tendency. On the contrary, different insights were achievable when applying MWPCA with a window size of 12 to dataset B (each line of the matrix is the average spectrum of the five experimental replicates). Different window sizes were tested for the different datasets, and the one selected was chosen since it allowed to achieve shapes for the dissimilarity index that could fit, according to the literature [29,39], with the process mechanism.

After having determined the dissimilarity index by observing their trends, three phases were identified and validated based on pH values (above 5.2, between 5.2 and 4.6 and below 4.6). In particular, the three phases were identified with pH values tendencies in order to obtain a reference for their occurrence. Subsequently, the phases were identified along the dissimilarity index plot according to the time of fermentation occurrence. The loadings variances calculated for each phase were also obtained. The interpretation allows to identify the information contained in each phase as well as the validation of the phase distinction. For example, Fig. 5 represents the dissimilarity index of batch #2 and the corresponding loading variance on PC1 and PC2 identified from different phases. The PCA model constructed within the first time-width is considered as reference, so the corresponding loadings as $\mathbf{p}(0)$. Since the statistical tool allows interpretation, raw reflectance data were initially used. Subsequently, data were converted into pseudo-absorbance values using the formula: $A = -\log(R)$. In addition, both reflectance and pseudo-absorbance spectra were smoothed with the Savitzsky Golay filter ($w = 7$, $p = 0$, $d = 0$) to identify the peak in the loading variance better. As a noticeable result, the dissimilarity index trends showed more evident tendencies by using reflectance data, while modeling the absorbance data allowed for better peak identification and interpretation.

By looking at the dissimilarity index, it is possible to state that the information did not lie only in one PC dimension since the trend over time could also be identified on PC2. The interpretation of the loadings variance in each phase permits learning about changes in the spectra that could be related to the kefir fermentation. For example, during phase 1, on PC 1, peaks at 1377, 1689, and 2207 emerged. All these changes could be attributed to the water hydration of charged structures and correspond to the existence of the liquid phase [29]. On the second

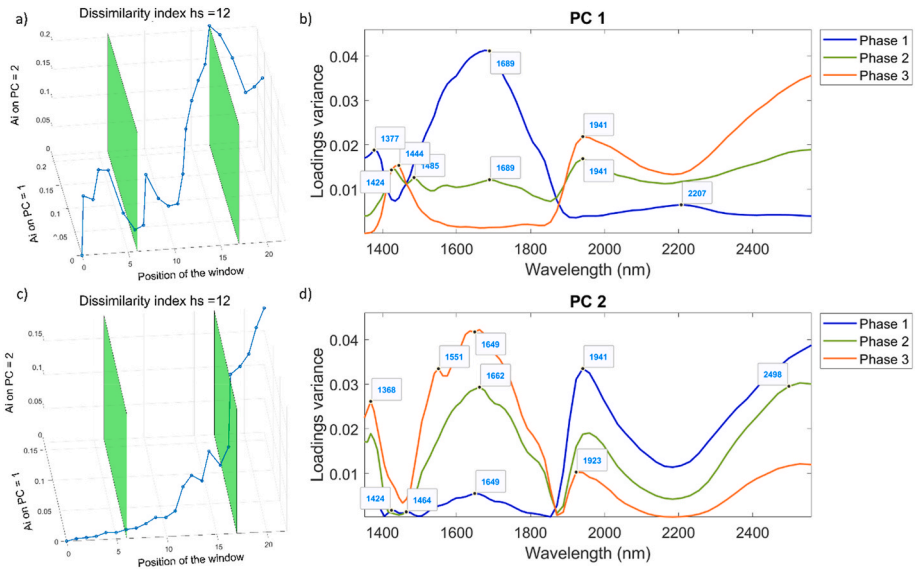

Fig. 5. Results of MW-PCA on batch #2, conducted at 25 °C and acquired after sample pretreatment. Window width = 12 and phase indexes identified as 1-6, 7-17, 18-22: a) dissimilarity index on PC1 and PC2 calculated for absorbance smoothed data, b) loading variance in the identified phases on PC1 (absorbance data), c) dissimilarity index on PC1 and PC2 calculated for reflectance smoothed data, d) loading variance in the identified phases on PC2 (absorbance data). The green planes show the indexes used to separate between phases. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Principal Component, an intense loading peak at 1941 nm, another critical region for O–H bonds emerged during the same phase. Another interesting result is that during the second phase, colored in dark green, on PC1, the more evident information is related to scattering that could be identified by the high offset of the loadings variance. Some information related to the water peaks is also identifiable. The results could be ascribable to the tumultuous phase of the fermentation, which is the most influenced by the change of viscosity. A general result is that using the miniaturized NIR data and looking into the changes in water structures makes it possible to gain insights into fermentations [44]. The chemical and physical changes that influence the spectra are more easily identified and separated by looking into the data with this tool, that do not assume the stationarity of the data.

All the fermentations were analyzed. Depending on the fermentation and sample pretreatment conditions, the phases needed to be slightly changed in accordance with the different biological times of the processes [41]. Identifying a suitable time window and the phase intervals is a key aspect of the method and was not trivial, even due to the noisy dissimilarity curves obtained. After the parameters were set, different pretreatment conditions led to different dissimilarity index trends and interesting peaks when observing the loadings variance. In addition, a notable variability between fermentation replicates was obtained (data not shown). The calculation was possible for all batches, and the interpretation of the phases' loadings could allow understanding if something unexpected is occurring. For example, in Supplementary Figs. S9, S10, S11, the dissimilarity index and the loadings variance for some fermentation conducted under different conditions are reported. The dissimilarity indexes for control batches were also calculated. The time intervals were set as the once for the homogenized fermentations at the specific temperature. From the loadings variance only one peak emerged in all phases, 1424 nm at 25 °C and 1444 nm at 30 °C.

In this context, MW-PCA allows the interpretation of data information, and it proved to be helpful for process understanding. It is worth mentioning that peak attribution could not be so immediate due to the broadness of NIR bands. For what concern process monitoring or endpoint detection, MW-PCA seems promising, allowing for interpretation and calibration-free identification of the fermentation state. As the first phases at 30 °C resulted short in the dissimilarity index plots (Supplementary Figs. S9 and S11), one of the main characteristics of this method emerged. A high number of time samplings could be needed for a suitable definition of the dissimilarity index and to gain reliable information.

3.3. *Paving the way for multivariate monitoring with process understanding: batches as array*

The 3D nature of the data acquired during experimentations was considered at this point. Data were unfolded into matrices in batch-wise and variable-wise directions (Supplementary Fig. S4). Typically, depending on the analysis goal, the unfolding type provides different information types. In batch-wise unfolding, the analysis focuses on the time-varying behavior over the batch duration, while in the variable-wise unfolding on the variable-wise changes. Camacho et al. widely discussed the advantages and disadvantages of the approaches [45].

3.3.1. *Batch-wise unfolding*

PCA was carried out on batch-wise unfolded reflectance and pseudo-absorbance data. Since the typical variability within batches of the data collected in this study is one of the relevant variances to consider in the models the use of a specific signal preprocessing for batches alignment was excluded. Score plots with different combinations of PC (from PC1 up to PC 4) were investigated and colored according to the temperature of the process and sample pretreatment. A total variance of about 92% was obtained both for reflectance and pseudo-absorbance data. PCA model on pseudo-absorbance data did not enhance any characteristic related to the conditions of process or acquisition. On the contrary, reflectance data distinguish between sample pretreatments. A possible explanation resides in the importance of the contributions to the spectra related to physical characteristics (mainly scattering) of the samples. PCA scores and loadings for reflectance data are reported in Fig. 6.

Loadings interpretation is possible and provides information about the spectra modification over time for different acquisition conditions. Since each variable value is highly correlated to the nearest ones, the trends could seem repetitive and the information could be somehow redundant. It is unlikely to see a change in a variable that did not affect the neighbors in a similar way, as can be seen in Fig. 6 d.

An interesting aspect of using this unfolding and a subsequent PCA resides in the fact that, at a glance, it is possible to identify

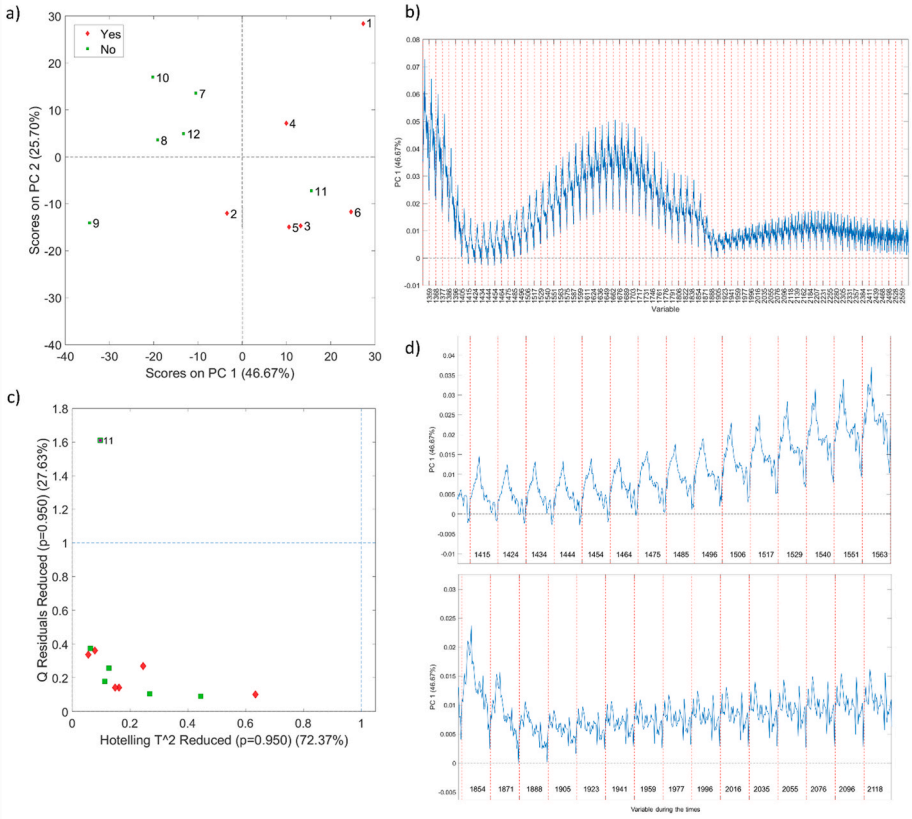

Fig. 6. Reflectance data. PCA batch-wise results: a) score plot colored by presence of sample pretreatment, b) loading plot with each variable ordered from time 1 to time 34, c) Influence plot derived from PCA with 2 PCs, d) zoom of the loading plot between 1415 and 1563 nm, and 1854 and 2118 nm.

abnormalities. Indeed, any outlier could be immediately identified by using the influence plot constructed with a suitable number of principal components (here two, according to eigenvalues and cumulative variance). As shown in Fig. 6 c, batch #11, which fell into the wrong group in the scores plot (Fig. 6 a), has a high Q residual, meaning that the fermentation profile caught by the spectra did not lie in the same variance structure as the other batches. In this case, if a deep investigation of the condition is needed, a Q contribution plot could be used to clarify the spectral regions responsible for that.

The advantage of this multivariate statistical tool is that an immediate and direct comparison of batches and their main differences are easily visualized and interpreted. Simple diagnostics could provide reasons for the differences. It could be handy for process understanding. After constructing a representative historical dataset, using the influence plot could be an effective tool for real-time process monitoring [45].

3.3.2. *Variable-wise unfolding*

After the variable-wise unfolding, different latent variable models were applied to extract all the information contained in the data.

3.3.2.1. PCA. First of all, PCA was applied. Scores and loadings are reported in Supplementary Fig. S12. Scores behavior was found to be really similar to that identified by PCA on individual X_i matrices. The advantage of this approach is the rapidity if more batches are included in the dataset and the possibility offered for direct comparison between scores trajectories for different batches. Indeed, by inspecting this Figure, insights are visible. The initial acquisition times are at negative values on PC1 for any fermentation. The scores values then increase progressively with different trajectories depending, first of all, on the sample pretreatment. An interesting aspect that emerged from scores visualization is the variability between the trajectories of different batches under the same conditions. Inspecting loadings allows for identifying the main common variable changes that characterize the process. Subsequently, as a test, control batches were included in the dataset and a new model was obtained and investigated. All the scores calculated fell at negative values with a normal distribution over time.

PCA on variable-wise unfolded data is a powerful technique for process understanding, even if, compared to MW-PCA, for example, it did not capture any dynamical information. Before efficiently implementing this PAT for process monitoring, an essential step could be evaluating the common and distinctive variance [46] related to the instrument [23,47] and the fermentation. The quantification and characterization of the uncertainty could be useful for developing ad hoc methods to remove or monitor even instrumental errors.

3.3.2.2. PLS for pH prediction. After exploiting the information in the spectra from different perspectives, a crucial aspect is determining the correlation between spectra and a parameter that could relate to the fermentation end-point. Some preliminary investigations for real-time process monitoring through PLS models were conducted. A property to correlate with spectra data must be identified. Since typically the end-point determination depends on the pH values as well as has been proposed for continuous monitoring, the values registered during the fermentation were used as a maturity index.

In the monitoring phase a fundamental aspect is determining what to intend as the normal operation conditions (NOC) and to assess the variability to include in the models. PLS models were calculated on the unfolded array composed by averaged spectra according to the different experimental conditions. Since the main goal of monitoring is to identify deviations from NOC, a global model including spectra from all the experimental condition was not considered here. For each PLS model leave-more-out (5 samples at time) Venetian Blind cross-validation was

performed and the number of latent variables was identified based on a good compromise between model complexity, and RMSEC, RMSECV, R^2 (CV) values.

Firstly, as calibration set two batches out of three for each condition were used to construct models sensitive to temperature and sample pretreatment. The excluded batch was used to validate the respective model. Batch#11 was used as validation set for batches at 30 °C with pretreatment of the samples. It was identified as an outlier from multiple strategies indeed, even in this case high residuals (both in influence plots and prediction capabilities of the model) were obtained. This represents a good example to enhance the potential for process monitoring of this statistical strategy.

Subsequently, depending on the sample pretreatment, four batches out of six were used to calibrate and two to validate (batch #11 was excluded to allow result comparison of the models' capabilities). In the end, data acquired under the same temperature conditions were joint. In Table 2, the performances of the PLS models are summed up.

Models with different complexities were obtained. The data investigation carried out in the previous paragraphs made it possible to infer that several variabilities were included in the data (e.g., spectra reproducibility over time, fermentation replicates, and fermentation biological times); therefore, rather complex models (high number of latent variables) were obtained. The cross-validation error ranged from 0.32 to 0.50, while the one in prediction was between 0.42 and 4.94. It is noticeable that the resulting errors obtained for homogenized batches are mostly higher than the errors for the unhomogenized batches. It is plausible that the introduction of an experimental step, namely the mixing process, increased the experimental variability present in the dataset, which subsequently manifested in the results of PLS models. The results obtained, even if preliminary, are promising under several perspectives for end-point determination. It has to be noticed that better estimation of the models could be obtained with a larger dataset as the variability between and within the batches could be better calculated. Indeed, by looking at predicted vs. measured plots, the need for a more extensive calibration set to identify the suitable model complexity and parameters better emerged. An example of the results obtained is reported in Supplementary Fig. S13. In any case, these preliminary results are useful to understand that specific models can be obtained to determine the end-point and monitor the process. In addition, from the results, it seemed reasonable to infer that models robust to slight temperature variations could also be obtained. The results for the models obtained by aggregating calibration data from pretreated and not pretreated samples at the respective temperatures (two last rows in Table 2) enhance the difficulties in joining these datasets.

In order to enhance the possibilities for monitoring, a PLS model was calibrated on all the spectra acquired with sample pretreatment. All the data acquired without stirring the samples were used as a validation set. The results obtained were RMSEC 0.38, RMSECV 0.43, RMSEP 1.29, R^2(CV) 0.609 and R^2(Pred) 0.003 with 6 LVs. The inspection of the

Table 2
Results for PLS models using different calibration and validation sets. NOC = normal operation conditions.

Data considered as NOC	#n of LVs	RMSEC	RMSECV	RMSEP	R^2 (CV)	R^2 (Pred)
Pretreated T = 25 °C	3	0.4	0.50	0.42	0.559	0.731
Pretreated T = 30 °C	6	0.26	0.37	4.94	0.739	0.151
No pretreatment T = 25 °C	4	0.27	0.32	0.72	0.747	0.589
No pretreatment T = 30 °C	5	0.22	0.39	0.56	0.710	0.532
Pretreated	6	0.32	0.43	0.78	0.630	0.580
No pretreatment	6	0.31	0.36	0.45	0.713	0.641
T = 25 °C	7	0.43	0.51	0.76	0.422	0.087
T = 30 °C	6	0.39	0.47	2.57	0.581	0.346

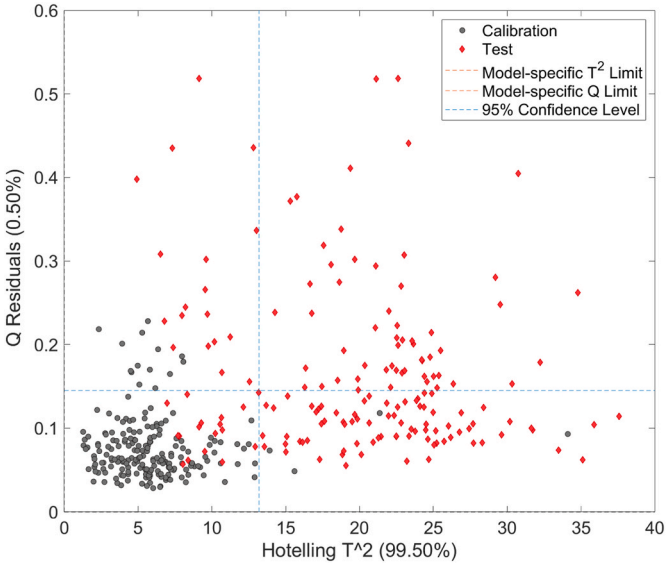

Fig. 7. Influence plot resulting from PLS variable-wise calibrated on batches from #1 to #6 and validated with batches from #7 to #12.

influence plot (Fig. 7) showed how faults in the homogenization system of the batches could be identified. Indeed, the test set mainly fell outside the T^2 and Q limits (identified with $\alpha = 95$ %), showing that those spectra did not follow the same covariance structure of the calibration set. Obviously, for monitoring as well as for effective end-point detection, a more extensive historical dataset would be needed to obtain implementable models.

4. Conclusions

A methodological approach has been proposed for evaluating miniaturized NIR spectrometers as PAT. Within this frame, after defining an acquisition strategy, it could be fundamental to understand what information is present in the data and to what extent it is available. The research provides results that could be used as a starting point for further implementations. In addition, an overview of the drawbacks and potentialities of statistical tools that are extensively used for benchtop instrumentation when there are limitations related to the different quality of data has been conducted.

This study proved the need for a combination of different multivariate statistical tools to exploit the information in NIR spectra acquired during a fermentation process. The complexity of the signals has been demystified through different perspectives enhancing the need for careful evaluation of miniaturized NIR data that could need a non-straightforward perspective. From the point of view of process understanding, each statistical technique brought out useful information regarding the process conditions.

Physical and chemical variations occurring in different phases of the process were identified, as well as differences under distinct process conditions were picked out. Homogenization of the fermentation could influence significantly the process and the spectra acquisition. Interesting insights about kefir transformation were also obtained: the main peaks and characteristics of spectra through the whole fermentation and different fermentation phases were identified. The influences of different temperatures were looked into for fermentations and spectra. It was found that higher temperatures increase the rate of the fermentation and that the information was contained also into the spectra. In the end, miniaturized NIR spectra were effectively demonstrated to include interesting information about kefir fermentation that allow to considered changes in the process conditions.

The monitoring perspective for abnormalities identification and end-point detection was considered. All the statistical tools allowed to

enhance some features related to the process and to understand the difficulties as, for instance, related to the heterogeneity of the medium. The potentialities from coupling NIR spectra and statistical tools were tested and proved.

CRediT authorship contribution statement

Gorla Giulia: Investigation, Formal analysis, Methodology, Conceptualization, Writing – original draft. **Alberto Ferrer:** Supervision, Conceptualization, Validation, Writing – review & editing. **Barbara Giussani:** Resources, Supervision, Conceptualization, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2023.341902.

References

[1] Y.Y. Pu, C. O'Donnell, J.T. Tobin, N. O'Shea, Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing, Int. Dairy J. 103 (2020), 104623, https://doi.org/10.1016/j.idairyj.2019.104623.

[2] D. Cozzolino, State-of-the-art advantages and drawbacks on the application of vibrational spectroscopy to monitor alcoholic fermentation (beer and wine), Appl. Spectrosc. Rev. 51 (2016) 282–297, https://doi.org/10.1080/05704928.2015.1132721.

[3] R.F. Sales, R. Vitale, S.M. de Lima, M.F. Pimentel, L. Stragevitch, A. Ferrer, Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy, Comput. Chem. Eng. 94 (2016) 343–353, https://doi.org/10.1016/j.compchemeng.2016.08.013.

[4] S. Grassi, A. Giraudo, C. Novara, N. Cavallini, F. Geobaldo, E. Casiraghi, F. Savorani, Monitoring chemical changes of coffee beans during roasting using real-time NIR spectroscopy and chemometrics, Food Anal. Methods (2023), https://doi.org/10.1007/s12161-023-02473-w.

[5] A.U. Vanarase, M. Alcalà, J.I. Jerez Rozo, F.J. Muzzio, R.J. Romañach, Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy, Chem. Eng. Sci. 65 (2010) 5728–5733, https://doi.org/10.1016/j.ces.2010.01.036.

[6] M.T. Islam, N. Scoutaris, M. Maniruzzaman, H.G. Moradiya, S.A. Halsey, M.S. A. Bradley, B.Z. Chowdhry, M.J. Snowden, D. Douroumis, Implementation of transmission NIR as a PAT tool for monitoring drug transformation during HME processing, Eur. J. Pharm. Biopharm. 96 (2015) 106–116, https://doi.org/10.1016/j.ejpb.2015.06.021.

[7] S. Wold, N. Kettaneh-Wold, J.F.F. MacGregor, K.G.G. Dunn, Batch process modeling and MSPC, in: Compr. Chemom. Chem. Biochem. Data Anal., VOLS 1-4, second ed., Elsevier, 2009, pp. A163–A197, https://doi.org/10.1016/b978-0-444-64165-6.03025-1.

[8] L. Rogers, K.F. Jensen, Continuous manufacturing-the green chemistry promise? Green Chem. 21 (2019) 3481–3498, https://doi.org/10.1039/c9gc00773c.

[9] F.M. de Andrade, R. Sales, N.C. da Silva, M.F. Pimentel, Calibration with virtual standards for monitoring biodiesel production using a miniature NIR spectrometer, Talanta 243 (2022), https://doi.org/10.1016/j.talanta.2022.123329.

[10] R.A. Crocombe, Portable spectroscopy, Appl. Spectrosc. 72 (2018) 1701–1751, https://doi.org/10.1177/0003702818809719.

[11] K.B. Beć, J. Grabska, C.W. Huck, Principles and applications of miniaturized near-infrared (NIR) spectrometers, Chem. Eur J. 27 (2021) 1514–1532, https://doi.org/10.1002/chem.202002838.

[12] K.B. Beć, J. Grabska, C.W. Huck, Miniaturized NIR spectroscopy in food analysis and quality control: promises, challenges, and perspectives, Foods 11 (2022) 1465, https://doi.org/10.3390/foods11101465.

[13] K.B. Beć, J. Grabska, H.W. Siesler, C.W. Huck, Handheld near-infrared spectrometers: where are we heading? NIR News 31 (2020) 28–35, https://doi.org/10.1177/0960336020916815.

[14] R. Zimmerleiter, J. Kager, R. Nikzad-Langerodi, V. Berezhinskiy, F. Westad, C. Herwig, M. Brandstetter, Probeless non-invasive near-infrared spectroscopic bioprocess monitoring using microspectrometer technology, Anal. Bioanal. Chem. 412 (2020) 2103–2109, https://doi.org/10.1007/s00216-019-02227-w.

[15] Z. Sun, C. Li, L. Li, L. Nie, Q. Dong, D. Li, L. Gao, H. Zang, Study on feasibility of determination of glucosamine content of fermentation process using a micro NIR spectrometer, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 201 (2018) 153–160, https://doi.org/10.1016/j.saa.2018.05.005.

[16] S. Grassi, C. Alamprese, V. Bono, C. Picozzi, R. Foschino, E. Casiraghi, Monitoring of lactic acid fermentation process using Fourier transform near infrared spectroscopy, J. Near Infrared Spectrosc. 21 (2013) 417–425, https://doi.org/10.1255/jnirs.1058.

[17] O. Arango, A.J. Trujillo, M. Castillo, Inline control of yoghurt fermentation process using a near infrared light backscatter sensor, J. Food Eng. 277 (2020) 1–6, https://doi.org/10.1016/j.jfoodeng.2019.109885.

[18] S. Grassi, L. Strani, E. Casiraghi, C. Alamprese, Control and monitoring of milk renneting using FT-NIR spectroscopy as a process analytical technology tool, Foods 8 (2019), https://doi.org/10.3390/foods8090405.

[19] C. Cimander, M. Carlsson, C.F. Mandenius, Sensor fusion for on-line monitoring of yoghurt fermentation, J. Biotechnol. 99 (2002) 237–248, https://doi.org/10.1016/S0168-1656(02)00213-4.

[20] M. Navrátil, C. Cimander, C.F. Mandenius, On-line multisensor monitoring of yogurt and filmjölk fermentations on production scale, J. Agric. Food Chem. 52 (2004) 415–420, https://doi.org/10.1021/jf0304876.

[21] A.L. Ntsame Affane, G.P. Fox, G.O. Sigge, M. Manley, T.J. Britz, Simultaneous prediction of acidity parameters (pH and titratable acidity) in Kefir using near infrared reflectance spectroscopy, Int. Dairy J. 21 (2011) 896–900, https://doi.org/10.1016/j.idairyj.2011.04.016.

[22] B. Giussani, G. Gorla, J. Riu, Analytical chemistry strategies in the use of miniaturised NIR instruments: an overview, Crit. Rev. Anal. Chem. (2022) 1–33, https://doi.org/10.1080/10408347.2022.2047607, 0.

[23] G. Gorla, P. Taborelli, C. Alamprese, S. Grassi, B. Giussani, On the importance of investigating data structure in miniaturized NIR spectroscopy measurements of food: the case study of sugar, Foods 12 (2023), https://doi.org/10.3390/foods12030493.

[24] G. Gorla, S. Fumagalli, J.J. Jansen, B. Giussani, Acquisition strategies for fermentation processes with a low-cost miniaturized nir-spectrometer from scratch: issues and challenges, SSRN Electron. J. 183 (2022), 108035, https://doi.org/10.2139/ssrn.4221663.

[25] M.L. Martínez, A. Garrido-Varo, E. De Pedro, L. Sánchez, Effect of sample heterogeneity on near infrared meat analysis: the use of the RMS statistic, J. Near Infrared Spectrosc. 6 (1998) 313–320, https://doi.org/10.1255/jnirs.214.

[26] E. Fernández-Ahumada, A. Garrido-Varo, J.E. Guerrero-Ginel, A. Wubbels, C. Van Der Sluis, J.M. Van Der Meer, Understanding factors affecting near infrared analysis of potato constituents, J. Near Infrared Spectrosc. 14 (2006) 27–35, https://doi.org/10.1255/jnirs.583.

[27] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37–52.

[28] S.S. Sekulic, H.W. Ward, D.R. Brannegan, E.D. Stanley, C.L. Evans, S.T. Sciavolino, P.A. Hailey, P.K. Aldridge, On-line monitoring of powder blend homogeneity by near-infrared spectroscopy, Anal. Chem. 68 (1996) 509–513, https://doi.org/10.1021/ac950964m.

[29] J. Muncan, K. Tei, R. Tsenkova, Real-time monitoring of yogurt fermentation process by aquaphotomics near-infrared spectroscopy, Sensors 21 (2021) 1–18, https://doi.org/10.3390/s21010177.

[30] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, A new multivariate statistical process monitoring method using principal component analysis, Comput. Chem. Eng. 25 (2001) 1103–1113, https://doi.org/10.1016/S0098-1354(01)00683-4.

[31] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (2014) 2812–2831, https://doi.org/10.1039/c3ay41907j.

[32] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.

[33] M.A. Farag, S.A. Jomaa, A.E.-W.E.-S. Aida, R. Hesham, The many faces of kefir fermented dairy products, Nutrients 12 (2020) 346. www.mdpi.com/journal.nutrients.

[34] A.E. Cervera, N. Petersen, A.E. Lantz, A. Larsen, K.V. Gernaey, Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation, Biotechnol. Prog. 25 (2009) 1561–1581, https://doi.org/10.1002/btpr.280.

[35] S. Treguier, C. Couderc, H. Tormo, D. Kleiber, C. Levasseur-Garcia, Identification of lactic acid bacteria Enterococcus and Lactococcus by near-infrared spectroscopy and multivariate classification, J. Microbiol. Methods 165 (2019), 105693, https://doi.org/10.1016/j.mimet.2019.105693.

[36] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta 1026 (2018) 8–36, https://doi.org/10.1016/j.aca.2018.04.004.

[37] J. Riu, G. Gorla, D. Chakif, R. Boqué, B. Giussani, Rapid analysis of milk using low-cost pocket-size NIR spectrometers and multivariate analysis, Foods 9 (2020) 1090, https://doi.org/10.3390/foods9081090.

[38] Z. Guzel-Seydim, J.T. Wyffels, A.C. Seydim, A.K. Greene, Turkish kefir and kefir grains: microbial enumeration and electron microscobic observation, Int. J. Dairy Technol. 58 (2005) 25–29, https://doi.org/10.1111/j.1471-0307.2005.00177.x.

[39] K. Zajšek, A. Goršek, Mathematical modelling of ethanol production by mixed kefir grains yeast population as a function of temperature variations, Biochem. Eng. J. 49 (2010) 7–12, https://doi.org/10.1016/j.bej.2009.11.004.

[40] J.A. Hageman, J.A. Westerhuis, A.K. Smilde, Temperature robust multivariate calibration : an overview of methods for dealing with temperature influences on near infrared spectra, J. Near Infrared Spectrosc. 62 (2005) 53–62.
[41] P. Jørgensen, J.G. Pedersen, E.P. Jensen, K.H. Esbensen, On-line batch fermentation process monitoring (NIR) - introducing "biological process time,", J. Chemom. 18 (2004) 81–91, https://doi.org/10.1002/cem.850.
[42] E.R. Farnworth, I. Mainville, Kefir - a fermented Milk product, in: Edward R. Farnworth (Ed.), Handb. Fermented Funct. Foods, second ed., second ed., CRC Press, 2008, pp. 89–128.
[43] C. Soukoulis, P. Panagiotidis, R. Koureli, C. Tzia, Industrial yogurt manufacture: monitoring of fermentation process and improvement of final product quality, J. Dairy Sci. 90 (2007) 2641–2654, https://doi.org/10.3168/jds.2006-802.
[44] R. Tsenkova, Aquaphotomics: acquiring spectra of various biological fluids of the same organism reveals the importance of water matrix absorbance coordinates and the aquaphotome for understanding biological phenomena, NIR News 19 (2008) 13–15, https://doi.org/10.1255/nirn.1060.
[45] J. Camacho, J. Picó, A. Ferrer, The best approaches in the on-line monitoring of batch processes based on PCA: does the modelling structure matter? Anal. Chim. Acta 642 (2009) 59–68, https://doi.org/10.1016/j.aca.2009.02.001.
[46] R. Vitale, O.E. de Noord, J.A. Westerhuis, A.K. Smilde, A. Ferrer, Divide et impera: how disentangling common and distinctive variability in multiset data analysis can aid industrial process troubleshooting and understanding, J. Chemom. 35 (2021) 1–12, https://doi.org/10.1002/cem.3266.
[47] G. Gorla, P. Taborelli, H.J. Ahmed, C. Alamprese, S. Grassi, R. Boqué, J. Riu, B. Giussani, Miniaturized NIR spectrometers in a nutshell: shining light over sources of variance, Chemosensors 11 (2023) 182, https://doi.org/10.3390/chemosensors11030182.