Food Control 191 (2027) 112478



Contents lists available at ScienceDirect

Food Control

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





Wine fermentation process control by PCA-based forecasting of infrared spectra

Jokin Ezenarro [a,*], Ricard Boqué [b], Daniel Schorn-García [b]

[a] Department of Food Science, University of Copenhagen, Frederiksberg, Denmark
[b] Universitat Rovira i Virgili, ChemoSens Group, Department of Analytical Chemistry and Organic Chemistry, Tarragona, Spain

ARTICLE INFO

Keywords:
Time series
Evolving window dissimilarity index
Process monitoring
MIR
Spectra prediction
Food

ABSTRACT

Early detection of fermentation deviations is essential, since delayed recognition of process disturbances can lead to quality losses, sluggish or stuck fermentations, and reduced production efficiency. This work extends Evolving Window Dissimilarity Index (EWDI)-based monitoring towards anticipatory process control by proposing a practical strategy to forecast EWDI from at-line ATR-FTIR spectra during wine fermentations. The workflow combines a moving-window PCA representation of the fermentation state with short-horizon extrapolation of principal component scores, reconstruction of future spectra, and EWDI calculation on the forecasted trajectories. Forecast uncertainty is estimated from the score prediction model and propagated to EWDI by Monte Carlo resampling. A probability of being in control (P_C) is also calculated from the forecasted EWDI distribution to express how likely the process is to remain within the control region.

Validation on under-control and intentionally deviated fermentations showed that EWDI can be forecast reliably in the short term, particularly in the early fermentation stage, enabling earlier recognition of emerging deviations. Longer horizons may diverge from the measured trajectory during phases of rapid kinetic change, where abrupt transitions cannot be anticipated until new measurements enter the calibration window. P_C provided a useful complement by distinguishing between predictions clearly outside the control region and predictions that, although centred within control limits, still showed enough uncertainty to indicate increased process risk. Together, EWDI forecasting and P_C provide a short-horizon, risk-oriented tool for fermentation monitoring.

1. Introduction

Industrial fermentation is a high-value process where small deviations can translate into significant economic losses through reduced yield, delayed tank turnover, reprocessing, quality downgrades or even lost batches. In winemaking, alcoholic fermentation is particularly sensitive to changes in temperature, nutrient availability and microbial dynamics; sluggish or stuck fermentations remain a practical concern because they can compromise both production planning and final product quality (Urtubia et al., 2008). For these reasons, there is sustained interest in process monitoring strategies that can detect deviations early enough to enable corrective actions while the process is still recoverable (Schorn-García et al., 2024).

Process Analytical Technology (PAT) frameworks promote the use of timely, in-process measurements to understand and control manufacturing processes, with the ultimate aim of building quality into the process rather than testing it at the end (U.S. Department of Health and Human Services, 2004). In the context of fermented beverages, infrared (IR) spectroscopy has become a particularly attractive PAT tool because it is rapid, non-destructive, reagent-free, and can be deployed at-line or on-line with relatively low running costs (Narayana et al., 2019). It also captures, in a single measurement, multivariate information associated with major compositional changes (e.g., sugar consumption, ethanol formation, organic acids), enabling real-time chemometric interpretation (Cozzolino, 2022). The suitability of mid-infrared (MIR) (Lambrecht et al., 2024), together with other spectroscopic techniques, such as visible-near-infrared (Vis-NIR) spectroscopy (Cozzolino et al., 2006) or Raman spectroscopy (Schorn-García et al., 2023) for monitoring wine fermentations and their quality has been demonstrated in multiple studies (Basalekou, 2025). A detailed tutorial specifically addressing portable ATR-FT-MIR as a PAT approach for wine alcoholic fermentation further supports the practical feasibility

* Corresponding author.
E-mail address: jokin@food.ku.dk (J. Ezenarro).

https://doi.org/10.1016/j.foodcont.2026.112478
Received 14 April 2026; Received in revised form 4 July 2026; Accepted 15 July 2026
Available online 17 July 2026
0956-7135/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

of at-line MIR monitoring in winery conditions (Schorn-García et al., 2021).

Most multivariate statistical process control (MSPC) approaches rely on projecting new observations onto a reference model built under normal operating conditions, and then tracking statistics such as Hotelling's T^2 and Q residuals (Catelani et al., 2018; Cavaglia et al., 2022). These PCA-based MSPC ideas are well established and widely used across process industries, including batch processes (Nomikos & MacGregor, 1995). However, an important difficulty arises when the deviation is primarily kinetic (a change in the speed of the trajectory) rather than a change in composition such as the spoilage caused by lactic acid bacteria (Cavaglia et al., 2020, 2022). In the first case, the process may remain chemically "on-path" but becomes temporally misaligned (slower or faster), which can reduce the sensitivity of conventional MSPC statistics when they are compared point-by-point in time (Schorn-García et al., 2024). This limitation has been highlighted in the context of wine fermentations, where deviations affecting kinetics can yield similar spectral features yet occur at different times, weakening the effectiveness of classical Q and T^2 charts for early detection. To address this specific challenge, the Evolving Window Dissimilarity Index (EWDI) was introduced as a monitoring statistic based on how the dominant PCA loading evolves as the process does (Schorn-García et al., 2024).

While this provides a practical solution for detection, modern process management increasingly requires moving beyond monitoring towards control. Once a deviation can be detected, the real value lies in anticipating its short-term evolution so that corrective actions (e.g., temperature adjustment, nutrient additions) can be applied promptly and proportionately. This would help minimise the impact of the deviation on the process and, ultimately, on the final product, including attributes associated with the synthesis and retention of volatile compounds. Forecasting within the framework of fermentation monitoring has therefore attracted attention, including early contributions linking statistical process monitoring concepts with operational decision making (Grigg & Walls, 2007). In chemometrics, forecasting is typically performed either by modelling the time evolution of latent variables (e.g., PCA scores) and predicting their trajectories, or by predicting spectral trajectories directly and subsequently deriving monitoring statistics from the forecasts (Esbensen & Swarbrick, 2018). While the latter approach can be sensitive to assumptions about spectral evolution and system variance (Feudale et al., 2002; Nomikos & MacGregor, 1994), the former approach is quite attractive, as it operates in a reduced latent space while enabling short-horizon predictions using low-dimensional time-series models applied to PCA scores, which are computationally efficient and suitable for routine monitoring applications. In addition, batch-process monitoring literature includes adaptive and multiway PCA strategies aimed at handling evolving trajectories without requiring strict synchronization, reflecting the broader need to accommodate non-stationarity and time-warping effects (Camacho & Picó, 2006; Lee et al., 2005; Rännar et al., 1998).

In this work, a fermentation control framework is proposed building on EWDI as a kinetics-sensitive monitoring statistic. We propose an approach to forecast EWDI by predicting short-term evolution in PCA score space from a moving window, reconstructing the corresponding future spectra, and propagating regression-based uncertainty through Monte Carlo simulation. The resulting framework provides not only detection of early deviations but also short-horizon forecasts with uncertainty bands, supporting timely and risk-aware process interventions in wine alcoholic fermentation monitored by IR spectroscopy.

2. Material and methods

2.1. Fermentation set-up

Small-scale alcoholic fermentations (microfermentations) were performed using a commercial concentrated white grape must (Juan Soler S.A., Cuenca, Spain) diluted with Milli-Q water to a final sugar concentration of 200 g L^{-1}, as described in (Schorn-García et al., 2024). For each microfermentation, 1.5 L of diluted must was transferred into 2 L glass vessels. Fermentations were initiated by inoculating the must with $3{\cdot}10^6$ CFU mL^{-1} of *Saccharomyces cerevisiae* (Viniferm Revelación, Agrovin S.A., Ciudad Real, Spain), following the manufacturer's rehydration instructions.

To generate both normal operation conditions (NOC) and controlled deviations, fermentations were run under different initial yeast assimilable nitrogen (YAN) conditions and temperature profiles, Supplementary Table S1. For NOC, TEM1, and TEM2, the YAN was adjusted, whereas the nitrogen-deficient fermentations (NDef) were prepared without this supplementation. Temperature was kept constant at 20 °C for NOC and NDef fermentations, whereas for TEM1 and TEM2, two different temperature gradient profiles were applied to create deviations. The NDT fermentation has both nitrogen deficit and temperature gradient 2.

Two experimental batches were conducted: (i) 6 NOC, 4 NDef, and 4 TEM1 fermentations; and (ii) 6 NOC, 4 TEM2, and 4 NDT fermentations. Fermentations within each set were carried out in parallel, with random preparation order within the set. Process evolution was monitored periodically by density and pH measurements using a portable densimeter (Densito2Go, Mettler Toledo, USA) and a portable pH meter with a 201 T electrode (7+series, XS Instruments, Italy); both instruments were calibrated with standards. Fermentation was considered finished when density fell below 0.995 g L^{-1}, typically reached after 7–9 days depending on the condition.

2.2. Spectroscopy

Mid-infrared spectra were acquired following a previously established at-line ATR-FTIR protocol using a portable 4100 ExoScan FTIR spectrometer (Agilent, California, USA) equipped with a spherical ATR interface and a diamond crystal (Schorn-García et al., 2024). Spectra were collected over 4000–650 cm^{-1} at 8 cm^{-1} resolution with 32 averaged scans, resulting in 899 measured wavenumbers (Fig. 1a). To reduce environmental contributions, spectra were corrected by an air-background spectrum acquired before each sample measurement.

Before sampling, each fermentation vessel was homogenised by gentle orbital stirring. Then, 1.5 mL of sample was collected and centrifuged at 10,000 rpm for 10 min to minimise scattering effects from microorganisms. Each sample was measured in triplicate by placing three successive drops on the ATR crystal; the three spectra were averaged for subsequent analysis. Sampling was performed at consistent times during the day, covering 12 daily hours of the process with a nominal sampling interval of 4 h. To better characterise the temperature gradients, higher-frequency sampling (one sample per hour) was performed for NOC, TEM1, TEM2, and NDT during the relevant periods.

Prior to chemometric analysis, spectra were preprocessed using a Savitzky–Golay second derivative (15-point window, second-order polynomial), followed by standard normal variate (SNV) correction (Fig. 1b), removing the extremes of the spectra (keeping from 3975 to 825 cm^{-1}, 845 variables in total) and mean centring (relative to the different subsets used in each case), as optimised by (Schorn-García et al., 2024).

2.3. Evolving Window Dissimilarity Index (EWDI)

The Evolving Window Dissimilarity Index (EWDI) is used as the monitoring statistic to quantify process deviations from a reference state using infrared spectra and PCA (Schorn-García et al., 2024), based on (Muncan et al., 2020). The core idea is that, if a given process has not started or is stable, the dominant spectral direction (captured by the first PCA loading) should remain similar to that of an initial reference period. Conversely, when the process is going on, the shape of the first loading changes; EWDI measures this change and the change rate as a dissimilarity index.

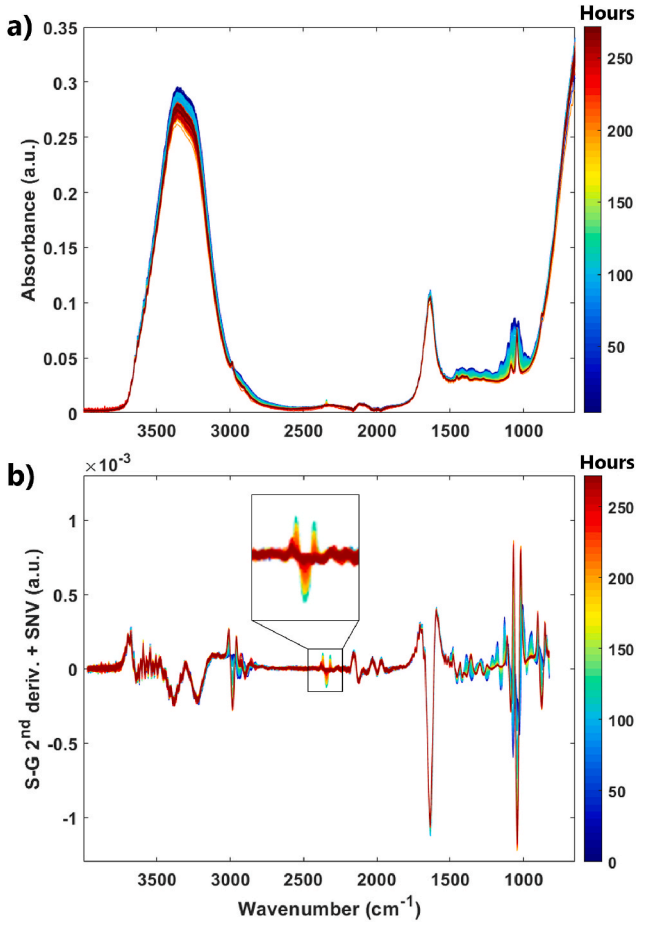


Fig. 1. ATR-FTIR spectral evolution during fermentations under control, colour-coded by sampling time in hours. a) Raw mid-infrared absorbance spectra. b) Corresponding Savitzky–Golay second derivative (15-point window, 2nd order polynomial) and SNV corrected spectra.

A reference PCA model ("Model 0") is first defined using an initial set of spectra acquired under nominal conditions (e.g., an early stable period of 48h), see Supplementary Fig. S1. Let $\mathbf{p}_0$ denote the first loading vector (PC1 loading) obtained from this reference model. Then, for each process time h, a PCA model is recomputed on an evolving data matrix that includes all spectra from the start up to h (i.e., an evolving window), see Fig. 2a. Let $\mathbf{p}_h$ denote the first loading of this evolving-window PCA model at time h (de Oliveira et al., 2017). Because PCA loadings are sign-indeterminate, similarity is computed via the absolute inner product (Equation (1)). Hence, the EWDI is defined as:

$$\mathrm{EWDI}(h) = 1 - \left| \mathbf{p}_h^{\mathrm{T}} \mathbf{p}_0 \right| \qquad \text{(Eq. 1)}$$

By construction, EWDI(h) takes values in [0, 1], this is, one minus the cosine of the angles between the two considered loading vectors. Values close to 0 indicate that the dominant spectral direction up to time h remains aligned with the reference direction (high similarity), whereas increasing values indicate progressive reorientation of the dominant spectral pattern relative to the reference, consistent with a process going on. Different types of process deviations change the rate at which the EWDI evolves over time, indicating that the process is going out of control (Schorn-García et al., 2024).

2.4. Forecasting the EWDI and uncertainty propagation

This section describes the procedure used to forecast short-term spectral evolution from a moving time window and to propagate the resulting predictive uncertainty to the downstream monitoring metric

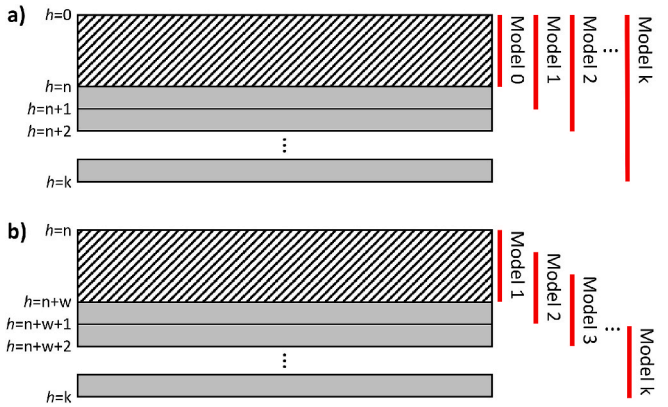


Fig. 2. Evolving vs. moving window strategies for PCA. a) Evolving (expanding) window: at each new time (h), newly acquired spectra (grey) are appended and the PCA model is rebuilt on all available data, so the model window grows with time. b) Moving (fixed-length) window: the PCA model is rebuilt on a constant-size window (w) that advances with time (older spectra are discarded as new ones enter), so the window length remains constant while its position shifts. Adapted from Schorn-García et al. (Schorn-García et al., 2024).

(EWDI, defined in Section 2.3). The approach operates in a low-dimensional latent space (PC scores) to obtain stable forecasts and an explicit uncertainty model (Fig. 3).

2.4.1. Forecasting the next spectra using the PC score space

At each update time, a calibration window of fixed duration (e.g., 48 h) is defined as the most recent subset of spectra available up to the current time; this is, a moving window; see Fig. 2b. Spectra in this window are mean-centred and a PCA model is computed (Equation (2)).

$$\mathbf{X}_w - \mathbf{1}\bar{\mathbf{x}}_w = \mathbf{T}_w \mathbf{P}_w^{\mathrm{T}} + \mathbf{E}_w \qquad \text{(Eq. 2)}$$

where $\mathbf{X}_w$ is the matrix of spectra in the calibration window, $\bar{\mathbf{x}}_w$ is their mean, $\mathbf{T}_w$ and $\mathbf{P}_w$ are the scores and the loadings for this window,

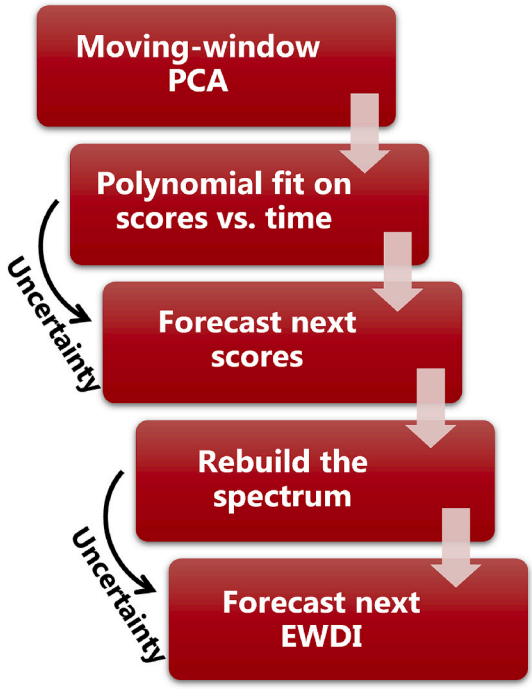


Fig. 3. Illustration of the workflow for EWDI forecasting and uncertainty propagation.

respectively, and $\mathbf{E}_w$ is the residual matrix. From this decomposition, the first K principal components (here $K = 3$) are retained to capture the dominant spectral evolution, representing the major sources of variance, but without including noise. Then, for each retained PC, its score values within the calibration window are modelled as a function of time using a low-order polynomial. For the k-th PC score t at hour h, the second order polynomial model (for instance) is:

$$\hat{t}_k(h) = \beta_{0,k} + \beta_{1,k}h + \beta_{2,k}h^2 \quad \text{(Eq. 3)}$$

The regression coefficients are estimated by ordinary least squares using only the time points and scores within the calibration window. The fitted models are then used to predict a future time vector defined by the prediction horizon and step (e.g., 10 time-points spaced by 4 h) to obtain the forecasts $\hat{\mathbf{t}}_k$, for $k = 1, \ldots, K$. Then, the spectra corresponding to those time-points ($\hat{\mathbf{x}}$ (h)) are rebuilt using the predicted scores and corresponding loadings from the training window PCA:

$$\hat{\mathbf{x}}(h) = \bar{\mathbf{x}}_w + \hat{\mathbf{t}}(h)\mathbf{P}_w^{\mathrm{T}} \quad \text{(Eq. 4)}$$

Then, with these forecasted spectra, the EWDI is calculated as proposed in the original article (Schorn-García et al., 2024), and explained in Section 2.3.

2.4.2. Uncertainty propagation

Uncertainty is propagated through the two steps of the EWDI forecasting: (i) the forecasting of the next scores and rebuilding the spectra using the PCA space, and (ii) the calculation of the EWDI from the forecasted spectra using Monte Carlo (MC) resampling.

For the first step, Equation (3) provides a mean forecast, but its standard deviation (square root of its variance) can also be estimated by using the following equation (Weisberg, 1985):

$$Var(t_k(h)) = \sigma_k^2 \cdot \left(1 + \frac{1}{n} + \frac{(h - \bar{h}_c)^2}{\sum (h_i - \bar{h}_c)^2}\right) = \sigma_k^2 \cdot (1 + \ell(h)) \quad \text{(Eq. 5)}$$

where σ_k^2 is the residual variance of the regression for the k-th PC estimated from the calibration window, $\boldsymbol{h}_c$ is the time-points used in calibration, h_i each of those time-points, n is the number of points, h is the future time-point being predicted and $\ell(h)$ is its leverage in the regression. The corresponding standard deviation is the square root of the variance. Then, based on the already obtained mean score forecasts across PCs, $\hat{\mathbf{t}}_k$ (with $K = 3$ in this work), and the corresponding deviations $\boldsymbol{s}_k$, a K-dimensional Gaussian distribution is defined. For propagating this uncertainty to the predicted EWDI values (EWDI*), N_{MC} random score vectors are drawn from this distribution, mapped back to spectra using the PCA basis of the calibration window (mean spectrum and loadings), and finally the EWDI values are computed as described in Section 2.3. The resulting Monte Carlo distribution of EWDI* values at each predicted point h is again described by its mean and standard deviation, and uncertainty bands are reported as:

$$\mu_{EWDI^*}(h) \pm z_{\alpha/2} \cdot \sigma_{EWDI^*}(h) \quad \text{(Eq. 6)}$$

where $\mu_{EWDI}(h)$ is the mean of the forecast population obtained by Monte Carlo resampling for time-point h, $\sigma_{EWDI}(h)$ is the standard deviation and $z_{\alpha/2}$ is the critical value of the standard normal distribution for a given significance level α ($z \approx 2$ for $\alpha = 0.05$). This approach yields EWDI forecasts accompanied by uncertainty estimates that are directly traceable to the regression-based uncertainty of the underlying score forecasts and that naturally increase with extrapolation through the leverage term.

2.5. Probability of being in control

To complement the EWDI forecast and its uncertainty band, a probability of being in control (P_C) is calculated for each forecasted future time point. The aim of this metric is to express, in a single value, what is the probability that the expected EWDI remains within the normal operating region defined by the control fermentations. For each future time point, the control region is described by the distribution of in-control EWDI curves, and its derived control limits (EWDI $\pm$ 2σ): the lower control limit (LLC) and the upper control limit (ULC). The forecasted EWDI* for the new fermentation is also represented by a Gaussian distribution, using the mean and standard deviation obtained from the Monte Carlo propagation of uncertainty (Fig. 4).

Based on this, the probability of being in control is then defined as the proportion of the forecasted EWDI* distribution falling inside the control limits at that same time point. In other words, P_C measures how much of the predicted uncertainty remains within the region expected for normal-operation fermentations. Mathematically, this can be written as:

$$P_C = P(LLC \leq EWDI^* \leq UCL) \cdot 100\% \quad \text{(Eq. 7)}$$

Values of P_C close to 100% indicate that the expected EWDI has a high probability of remaining within the control region, whereas values close to 0% indicate that the forecast is unlikely to be compatible with normal operation.

2.6. Software

All data processing, modelling, forecasting, and uncertainty propagation were implemented in the MATLAB environment (v. 2024a, The MathWorks, Natick, MA, USA) using custom-written scripts and functions. The full workflow, including the routines for moving-window PCA, score forecasting, Monte Carlo uncertainty propagation, and EWDI computation, is provided as open-source code in a public GitHub repository: https://github.com/JokinEzenarro/FEWDI/

3. Results and discussion

3.1. Method optimisation

The forecasting approach described in Sections 2.3–2.4 requires selecting (i) the length of the moving calibration window used to fit the score–time model and (ii) the order of the polynomial fitted to the score values, as these will affect forecast horizon over which predicted EWDI values remain reliable. To avoid information leakage and keep an independent validation set, the optimisation was performed exclusively on Batch 2 control fermentations, while results reported in the following sections describe control fermentations from Batch 1.

A grid search optimisation was conducted by varying the training

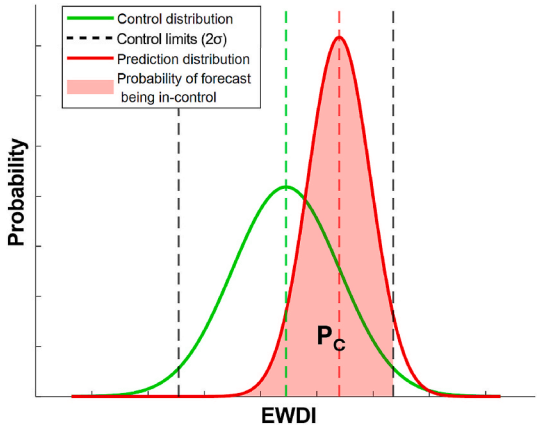


Fig. 4. Illustration of the control EWDI distribution at a certain time point (green) and the predicted EWDI and the corresponding uncertainty distribution (red) for another process, where the probability of the forecast to stay in control (P_C) is shaded in red; 95.8% in the example.

window length (12–60 h) and polynomial order (1st or 2nd), while evaluating prediction performance as a function of the forecast step (first to tenth future time-point). For each configuration, EWDI was predicted from forecast spectra and compared with the EWDI computed from the measured spectra at the corresponding future times. Meaning that instead of predicting a moving fixed-hour window (as it in the calibration window), the exact time-points where a real measurement lies were predicted, so each predicted time-point matches with one of the next measured EWDI values, and they can be compared. Performance was summarised using the correlation coefficient R (Fig. 5).

Two clear trends were observed. First, prediction accuracy generally decreased with increasing forecast horizon, as expected. Second, the training window length modulated both stability and robustness: very short windows tended to produce more variable performance, whereas longer windows gave smoother and more consistent behaviour across prediction steps. The comparison between polynomial orders highlights an important trade-off. With small calibration windows, the second-order polynomial produced markedly poorer and less stable results, particularly for the earliest prediction steps (Fig. 5). This behaviour is consistent with overfitting when curvature is estimated from limited information, leading to unreliable extrapolation even at a short distance beyond the window. In contrast, as the window length increased, the second-order model became more competitive and improved longer-horizon behaviour. This is plausible because, over longer time windows, the fermentation trajectory may no longer be well approximated by a purely linear trend, so allowing curvature can better represent the underlying evolution. Taking this into account, the parameters selected as optimal were a 2nd order polynomial and a 36-h calibration window (Fig. 5). This configuration delivered consistently strong correlations across the early forecast points where the process evolves rapidly and practical decision-making is most relevant, while keeping a relatively narrow calibration window.

It must also be considered that the number of PCs kept in the moving window PCA also affects the performance. Even though the first PC explains around 95% of the process variance, it fails at modelling the subtle details that make the different processes indeed different. Three PCs explain around 99% of the variance, including the different particularities, while subsequent PCs mostly explain noise and as they are so low, they do not affect the performance; therefore, three PCs were used as optimal. Apart from this, it was also tested whether emphasising spectra closer in time to the prediction point improved performance. Both weighted PCA and a weighted linear regression were evaluated using weights proportional to the inverse time distance (more recent samples receive higher weights) (Van Deun et al., 2019). These weighted variants produced equivalent performance to the unweighted approach in the optimisation study; consequently, the simpler unweighted workflow was retained for the remainder of the study.

From an oenological perspective, the selection of a 36-h calibration window is also meaningful. In the first models, windows that are too short mainly reflect the initial adaptation phase of the inoculated yeast, when biomass establishment and metabolic adjustment still dominate the trajectory. By contrast, a 36-h window already includes the transition towards the tumultuous stage of alcoholic fermentation, where sugar consumption, ethanol formation, CO_2 release and heat production become more structured. This likely explains why the model benefits from a window long enough to predict the fermentative direction while remaining local enough to detect emerging deviations.

3.2. Forecasting the EWDI

After selecting the forecasting configuration using Batch 2, NOC fermentations from Batch 1 and the deviated fermentations are discussed here to ensure an independent validation and avoid information leakage. Fig. 6 shows EWDI forecasts for different fermentation conditions and for three time-points (50 h, 60 h and 100 h). The full set of forecasts across the entire process is provided in the Supplementary Videos S1-S5. Supplementary Fig. S2 shows an overview of the average EWDI profiles for these processes.

At 50 h, corresponding to the onset of the tumultuous phase of the fermentation and where the first EWDI values can be calculated, the method already provides useful short-horizon predictions, but it is important to note a practical detail of the implementation: at this early stage the calibration window is not yet long enough to reach the target minimum window length (36 h). Consequently, the PCA used for forecasting behaves as an evolving window until sufficient data are available; only after that point does it become a true moving window (see Fig. 2). This is reflected in a wider uncertainty and less stable extrapolation early in the process. This is oenologically reasonable because the process is then highly dynamic: sugar catabolism is accelerating, ethanol and CO_2 are accumulating rapidly, and the fermenting medium is undergoing major physicochemical changes (Ribéreau-Gayon et al., 2021). In this region, short-term forecasting (≈10 h) is generally reliable, whereas longer horizons (>20 h) can deviate substantially from the measured EWDI. However, the trustfulness (or lack of trust) for the

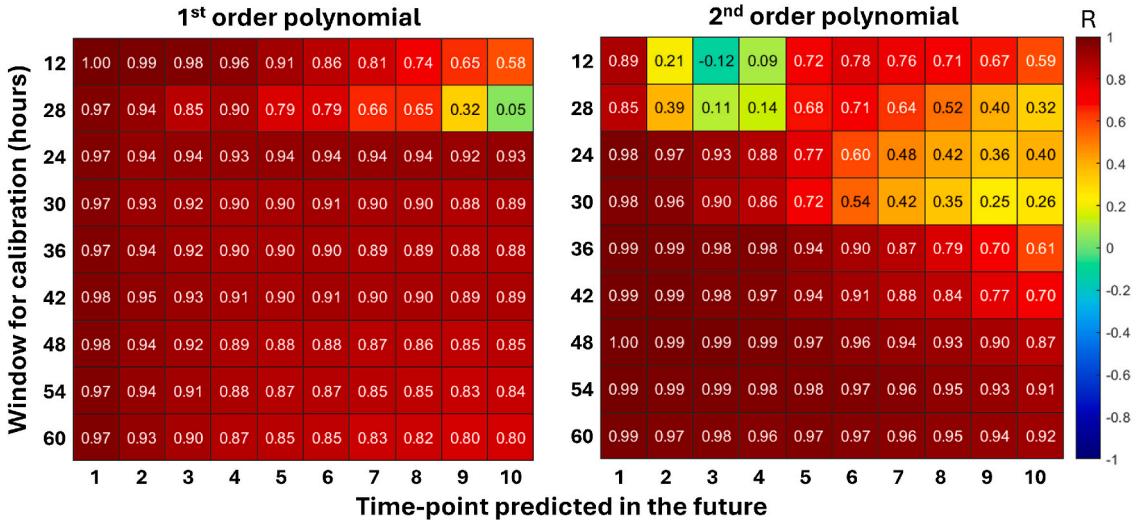


Fig. 5. Grid-search optimisation of the forecasting configuration using Batch 2. Heatmaps show the correlation coefficient (R) between predicted and reference EWDI as a function of the moving calibration window length (12–60 h, y-axis) and the forecast step (1–10 future time-points, x-axis), for polynomial models of different degree (1st and 2nd; left and right panels, respectively).

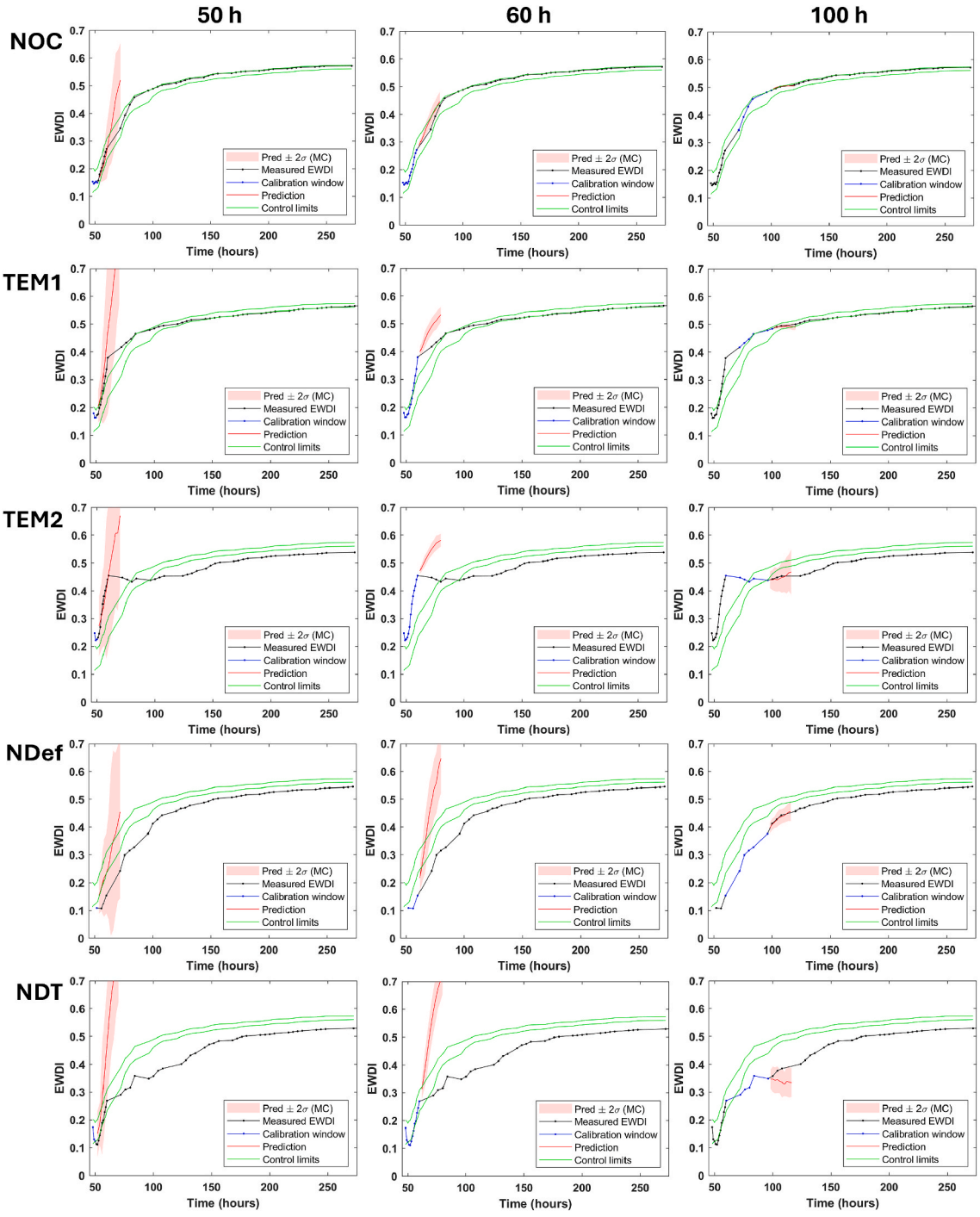


Fig. 6. EWDI forecasting examples across fermentation conditions and process stages. Panels show representative EWDI forecasts for five fermentation conditions (rows: NOC, TEM1, TEM2, NDef, NDT) at three different update times during the process (columns: 50, 60 and 100 h). The true EWDI trajectory is shown in black, the calibration (training) window used for forecasting is highlighted in blue, and the predicted EWDI segment is shown in red with its uncertainty band (shaded; mean ± 2σ, indicating a 95% confidence interval). Green curves indicate the control limits.

prediction is signalled by its uncertainty, indicating whether the EWDI is likely to remain within control limits or approach an out-of-control region; which sets a relative basis for early attention and planning of potential corrective actions.

At 60 h, forecasts tend to be accurate for control (NOC) fermentations, where EWDI evolves smoothly and the local score–time relationship remains somehow stable over the calibration window. Under these conditions, the predicted EWDI closely tracks the measured trajectory, and the uncertainty bands are comparatively tighter. In contrast, for out-of-control fermentations characterised by abrupt kinetic changes (subjected to temperature disturbances mimicking temperature control malfunctions: TEM and NDT mostly), the proposed approach fails to anticipate the turning behaviour: the forecast is based on the recent past within the calibration window, and therefore cannot predict a sudden change in speed that is not yet present in the measurements. This is also coherent from an oenological point of view: a moderate temperature rise may temporarily accelerate metabolism, but once the system moves beyond the most favourable range, yeast physiology becomes increasingly stressed, particularly in the presence of rising ethanol, and kinetics can change suddenly. Forecast performance is recovered once new

spectra enter the window and the updated scores reflect the abrupt transition, after which the model adapts and subsequent short-horizon predictions improve (see Supplementary Videos S1-S5). For nitrogen-deficient fermentations (NDef), the slower EWDI evolution is consistent with limited assimilable nitrogen restricting fermentative activity. This is relevant not only because it slows sugar depletion, but also because nitrogen status is known to modulate the formation of fermentation-derived volatiles and the risk of formation of reductive off-flavours (Torrea et al., 2011).

At 100 h and later, the fermentation trajectories become more stable and EWDI tends to approach a slower, near-plateau evolution across conditions. In this regime the score dynamics are smoother, the regression fit becomes more stable, and the forecasts are consistently accurate with low uncertainty. This stage corresponds to a more settled phase in which the major compositional turnover has already occurred, and the process trajectory becomes easier to extrapolate. This behaviour is visible in Fig. 6 and is further supported by the complete time-series forecasts in the Supplementary Videos S1-S5, where the later-stage predictions remain close to the measured EWDI and the uncertainty bands narrow accordingly.

It should also be noted that the reliability of the forecast is strongly dependent on the number of spectra available in the calibration window. When the window contained very few measurements (in practice, fewer than ~5 spectra), the polynomial fit of scores versus time becomes unstable and can produce erratic extrapolations, because the model is forced to infer a trajectory from insufficient information and is highly sensitive to small local fluctuations. As the number of spectra in the window increases, the score–time relationship is better described, the estimated trend becomes smoother, and both the forecast and its uncertainty behave more consistently. This is especially important in wine fermentation, where short periods of metabolic instability can have consequences disproportionate to their duration. This underlines a practical advantage of increasing sampling density: the proposed strategy would substantially benefit substantially from real-time or near real-time monitoring strategy, significantly increasing the available data for calibration and forecasting; particularly during rapidly changing phases of the fermentation (notably around ~60 h), where kinetics can shift quickly.

3.3. Probability of being in control

The interpretation of the forecasted EWDI values and their uncertainty can be complemented by the use of the P_C (probability of being in control) metric. Whereas the results in Section 3.2 were discussed in terms of forecasted EWDI trajectories and their associated uncertainty bands, P_C provides an additional probabilistic summary of the same information. Table 1 shows the P_C values obtained for the EWDI forecasts at hour 50 for the different fermentations. In this context, the main contribution of P_C is not to modify the interpretation already obtained from the EWDI forecasts, but to make it more explicit and quantitative, easier to compare across samples and forecast horizons. A high P_C value indicates that the forecasted EWDI, including its uncertainty, is still largely compatible with the control region. However, a low value does not necessarily indicate that the process will get out of control, it could be both that the forecasted EWDI fall well beyond the control limits or that the corresponding uncertainty is too big to ensure that it will remain under control.

Table 1
P_C values for each of the forecasted EWDI points at hour 50 for each of the fermentations.

P_C (%)	Timepoint forecasted			
	1h	2h	5h	10h
NOC	99.6%	84.4%	58.3%	23.4%
TEM1	44.9%	17.4%	14.0%	4.3%
TEM2	5.6%	4.5%	5.9%	4.7%
NDef	46.5%	45.7%	7.8%	1.0%
NDT	75.9%	74.9%	29.4%	2.3%

For the NOC fermentation, the high P_C values at the shortest forecast horizons confirm in probabilistic terms that the predicted EWDI remains strongly compatible with normal-operation behaviour. As the horizon increases, the decrease in P_C reflects the progressive reduction in confidence already observed in Section 3.2. In this way, P_C converts the combined effect of forecast deviation and uncertainty widening into a single, directly interpretable quantity.

For the deviated fermentations, the added value of P_C is particularly clear because it captures not only whether the forecasted EWDI is centred inside or outside the control region, but also how certain that prediction is. When the predicted EWDI distribution lies almost entirely outside the control limits, P_C becomes very low, clearly indicating an out-of-control forecast. However, a decrease in P_C may also occur in cases where the predicted mean EWDI still falls within the control region, but the associated uncertainty is large. In such situations, a substantial part of the predictive distribution extends beyond the control limits, reducing P_c and signalling that the process should still be considered at risk. In this way, P_C provides a more informative measure than visual inspection alone, because it reflects both deviation from control behaviour and the uncertainty associated with that prediction.

4. Conclusions

In this work, the EWDI-based process monitoring was extended towards anticipatory control by proposing a strategy to forecast EWDI from at-line ATR-FTIR data. The approach combines moving-window PCA to represent fermentation dynamics, short-horizon extrapolation of score trajectories, reconstruction of future spectra, and EWDI computation on predicted states. Uncertainty was estimated from the linear-regression prediction of scores and propagated to EWDI via Monte Carlo resampling, yielding confidence intervals linked to both calibration window and forecast horizon. The probability of being in control (P_C) complements the EWDI by summarising the predicted value and its uncertainty into a single probabilistic measure of compliance with the control region.

Validation on controlled and deviated fermentations showed reliable short-term EWDI forecasting (≈10 h) during early fermentation, enabling early detection of deviations. However, accuracy decreased at longer horizons, especially during rapid kinetic transitions (~60 h). Performance is highest when EWDI evolves smoothly, while abrupt deviations are only captured once updated data enter the calibration window. P_C improves interpretability by distinguishing between truly out-of-control predictions and high uncertainty around borderline states. In this sense, forecast quality strongly depends on calibration window content: sparse spectral sampling led to unstable trends, whereas denser data significantly improves both predictions and confidence estimates. This supports the need for real-time acquisition (through the implementation of in-line or on-line measurements), particularly during fast kinetic phases where early intervention is most valuable.

Overall, the proposed method enhances process surveillance and enables more proactive management of the wine alcoholic fermentation (and potentially other processes) purely based on rapid non-destructive measurements. This method was validated with oenologically relevant process deviations: temperature variations and nitrogen deficiency, which affect yeast metabolism, fermentation completion, and aroma formation. Short-horizon EWDI forecasting combined with P_C provided an early-warning tool to support corrective actions, aiming at mitigating the impact of those deviations on wine quality.

Funding

This work was funded by the Novo Nordisk Foundation grant number

NNF21OC0065495 and grant PID2024-156237OB-C31 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, EU.

CRediT authorship contribution statement

Jokin Ezenarro: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Ricard Boqué:** Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing. **Daniel Schorn-García:** Data curation, Investigation, Methodology, Validation, Visualization, 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.

Acknowledgements

During the preparation of this work, the authors used ChatGPT in order to improve the understandability and conciseness of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2026.112478.

Data availability

Data will be made available on request.

References

Basalekou, M. (2025). Using IR spectroscopy as a holistic monitoring approach in winemaking: A review. *OENO One, 59*(3). https://doi.org/10.20870/oeno-one.2025.59.3.7824

Camacho, J., & Picó, J. (2006). Online monitoring of batch processes using multi-phase principal component analysis. *Journal of Process Control, 16*(10), 1021–1035. https://doi.org/10.1016/j.jprocont.2006.07.005

Catelani, T. A., Santos, J. R., Páscoa, R. N. M. J., Pezza, L., Pezza, H. R., & Lopes, J. A. (2018). Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta, 179*, 292–299. https://doi.org/10.1016/j.talanta.2017.11.010

Cavaglia, J., Garcia, S. M., Roger, J.-M., Mestres, M., & Boqué, R. (2022). Detection of bacterial spoilage during wine alcoholic fermentation using ATR-MIR and MCR-ALS. *Food Control, 142*, Article 109269. https://doi.org/10.1016/J.FOODCONT.2022.109269

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

Cozzolino, D. (2022). Advantages, opportunities, and challenges of vibrational spectroscopy as tool to monitor sustainable food systems. *Food Analytical Methods, 15*(5), 1390–1396. https://doi.org/10.1007/s12161-021-02207-w

Cozzolino, D., Parker, M., Dambergs, R. G., Herderich, M., & Gishen, M. (2006). Chemometrics and visible-near infrared spectroscopic monitoring of red wine fermentation in a pilot scale. *Biotechnology and Bioengineering, 95*(6), 1101–1107. https://doi.org/10.1002/bit.21067

de Oliveira, R. R., Pedroza, R. H. P., Sousa, A. O., Lima, K. M. G., & de Juan, A. (2017). Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy. *Analytica Chimica Acta, 985*, 41–53. https://doi.org/10.1016/j.aca.2017.07.038

Esbensen, K. H., & Swarbrick, B. (2018). *Multivariate data analysis* (6th ed.) Camo.

Feudale, R. N., Woody, N. A., Tan, H., Myles, A. J., Brown, S. D., & Ferré, J. (2002). Transfer of multivariate calibration models: A review. *Chemometrics and Intelligent Laboratory Systems, 64*(2), 181–192. https://doi.org/10.1016/S0169-7439(02)00085-0

Grigg, N., & Walls, L. (2007). The role of control charts in promoting organisational learning: New perspectives from a food industry study. *The TQM Magazine, 19*(1), 37–49. https://doi.org/10.1108/09544780710720826

Lambrecht, K., Fonseca Diaz, V., Saeys, W., Louw, T., du Toit, W., & Aleixandre-Tudo, J. L. (2024). Forecasting wine phenolic composition from infrared spectra of grapes extracts and monitoring of fermentations with optimised time-specific prediction models. *Chemometrics and Intelligent Laboratory Systems, 248*, Article 105114. https://doi.org/10.1016/j.chemolab.2024.105114

Lee, D. S., Park, J. M., & Vanrolleghem, P. A. (2005). Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology, 116*(2), 195–210. https://doi.org/10.1016/j.jbiotec.2004.10.012

Muncan, J., Tei, K., & Tsenkova, R. (2020). Real-time monitoring of yogurt fermentation process by aquaphotomics near-infrared spectroscopy. *Sensors, 21*(1), 177. https://doi.org/10.3390/s21010177

Narayana, S., Christensen, L., Skov, T., & van den Berg, F. (2019). Mid-infrared spectroscopy and multivariate analysis to characterize Lactobacillus acidophilus fermentation processes. *Applied Spectroscopy, 73*(9), 1087–1098. https://doi.org/10.1177/0003702819848486

Nomikos, P., & MacGregor, J. F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE Journal, 40*(8), 1361–1375. https://doi.org/10.1002/aic.690400809

Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics, 37*(1), 41. https://doi.org/10.2307/1269152

Ribéreau-Gayon, P., Doneche, B., Lonvaud, A., & Darriet, P. (2021). *Handbook of enology. Volume 1, the microbiology of wine and vinifications*. Wiley.

Rännar, S., MacGregor, J. F., & Wold, S. (1998). Adaptive batch monitoring using hierarchical PCA. *Chemometrics and Intelligent Laboratory Systems, 41*(1), 73–81. https://doi.org/10.1016/S0169-7439(98)00024-0

Schorn-García, D., Cavaglia, J., Giussani, B., Busto, O., Aceña, L., Mestres, M., & Boqué, R. (2021). ATR-MIR spectroscopy as a process analytical technology in wine alcoholic fermentation-A tutorial. *Microchemical Journal, 166*, Article 106215. https://doi.org/10.1016/j.microc.2021.106215

Schorn-García, D., Ezenarro, J., Aceña, L., Busto, O., Boqué, R., Giussani, B., & Mestres, M. (2023). Spatially offset raman spectroscopic (SORS) analysis of wine alcoholic fermentation: A preliminary study. *Fermentation, 9*(2), 115. https://doi.org/10.3390/fermentation9020115

Schorn-García, D., Ezenarro, J., Busto, O., Aceña, L., Boqué, R., Mestres, M., & Giussani, B. (2024). A new index to detect process deviations using IR spectroscopy and chemometrics process tools. *Food and Bioprocess Technology, 17*(9), 2782–2792. https://doi.org/10.1007/s11947-023-03266-z

Torrea, D., Varela, C., Ugliano, M., Ancin-Azpilicueta, C., Leigh Francis, I., & Henschke, P. A. (2011). Comparison of inorganic and organic nitrogen supplementation of grape juice – Effect on volatile composition and aroma profile of a Chardonnay wine fermented with Saccharomyces cerevisiae yeast. *Food Chemistry, 127*(3), 1072–1083. https://doi.org/10.1016/j.foodchem.2011.01.092

Urtubia, A., Pérez-correa, J. R., Pizarro, F., & Agosin, E. (2008). Exploring the applicability of MIR spectroscopy to detect early indications of wine fermentation problems. *Food Control, 19*(4), 382–388. https://doi.org/10.1016/j.foodcont.2007.04.017

U.S. Department of Health and Human Services. (2004). *Guidance for industry PAT — A framework for innovative pharmaceutical development, manufacturing, and quality assurance*.

Van Deun, K., Thorrez, L., Coccia, M., Hasdemir, D., Westerhuis, J. A., Smilde, A. K., & Van Mechelen, I. (2019). Weighted sparse principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 195*, Article 103875. https://doi.org/10.1016/j.chemolab.2019.103875

Weisberg, S. (1985). *Applied linear regression*. Wiley.