| RESEARCH ARTICLE OPEN ACCESS

Raman-Enabled Predictions of Protein Content and Metabolites in Biopharmaceutical *Saccharomyces cerevisiae* Fermentations

Jeppe Hagedorn[1,2] 🄳 | Guilherme Ramos[3] | Miguel Ressurreição[3] | Ernst Broberg Hansen[4] | Michael Sokolov[5] | Carlos Casado Vázquez[6] | Christos Panos[7]

[1]SDU Chemical Engineering, University of Southern Denmark, Odense, Denmark | [2]Process Analytical Technology, Novo Nordisk A/S, Måløv, Denmark | [3]DataHow LDA, Lisboa, Portugal | [4]CMC API Development, Novo Nordisk A/S, Bagsværd, Denmark | [5]DataHow AG, Dübendorf, Switzerland | [6]Microbial Cultivation Development, Novo Nordisk A/S, Bagsværd, Denmark | [7]Chemical Development, Novo Nordisk A/S, Bagsværd, Denmark

Correspondence: Jeppe Hagedorn (jqhh@novonordisk.com)

ABSTRACT

Raman spectroscopy, a robust and non-invasive analytical method, has demonstrated significant potential for monitoring biopharmaceutical production processes. Its ability to provide detailed information about molecular vibrations makes it ideal for the detection and quantification of therapeutic proteins and critical control parameters in complex biopharmaceutical mixtures. However, its application in *Saccharomyces cerevisiae* fermentations has been hindered by the inherent strong fluorescence background from the cells. This fluorescence interferes with Raman signals, compromising spectral data accuracy. In this study, we present an approach that mitigates this issue by deploying Raman spectroscopy on cell-free media samples, combined with advanced chemometric modeling. This method enables accurate prediction of protein concentration and key process parameters, fundamental for the control and optimization of biopharmaceutical fermentation processes. Utilizing variable importance in projection (VIP) further enhances model robustness, leading to lower relative root mean squared error of prediction (RMSEP) values across the six targets studied. Our findings highlight the potential of Raman spectroscopy for real-time, on-line monitoring and control of complex microbial fermentations, thereby significantly enhancing the efficiency and quality of *S. cerevisiae*-based biopharmaceutical production.

1 | Introduction

Raman spectroscopy is a well-established technique within the field of process analytical technology (PAT) [1], valued for its insensitivity to water and non-invasive nature [2, 3]. Over the past three decades, Raman spectroscopy has been widely applied in bioprocessing [4–7] and biopharmaceutical industries [8–11]. The multivariate nature of Raman spectra necessitates the use of chemometrics for effective process monitoring and control [12, 13].

In the past 10–15 years, the use of Raman spectroscopy has increased, particularly within mammalian cell-based recombinant protein production. Comparatively, Raman spectroscopy has been used as a PAT for mammalian cultivation to a greater extent than for microbial fermentation [14]. This can be attributed

---

Summary

- This study presents a scalable method that offers the possibility of continuous monitoring compared to measuring discrete samples using traditional HPLC and other analytical methods currently employed in biopharmaceutical production.

- By utilizing Raman spectroscopy, our approach streamlines the assessment of critical process and quality parameters in *Saccharomyces cerevisiae* fermentations.

- Specifically developed for industrial use, this method showcases a novel application of Raman spectroscopy in yeast-based fermentations for biopharmaceutical protein production.

- Additionally, it holds the potential to broaden the scope of Raman spectroscopy in the fermentation-based production and processing of recombinant therapeutic peptides or proteins.

- By demonstrating this new application space, we aim to inspire further advancements in the use of Raman spectroscopy for fermentation analysis in the biopharmaceutical industry.

to several factors, which have made using Raman spectroscopy for process monitoring of mammalian cell cultivations more feasible than their microbial counterparts: (1) The lower biomass, even when operating in perfusion mode [14, 15]. The biomass dry weight for high-yielding mammalian cultivations is typically below 0.1% [16, 17], while for *Saccharomyces cerevisiae* it can reach upwards of 10% w/w or more. (2) The (auto)fluorescence background, which although high for both, is much lower for mammalian cells [18]. This can be attributed to the media composition [8, 19, 20], and the cell constituents, for example differences in cell wall structure [21] and metabolites [22].

1.1 | Current State of the Art for Monitoring Yeast Fermentations

Process monitoring of yeast fermentations is a critical aspect of biomanufacturing. The critical process parameters (CPP), which are optimized to produce the most stable high-yielding processes are: temperature, pH, oxygen supply, growth rate, and nutrient availability. The primary goal of process monitoring is to ensure consistent product quality and yield, while minimizing the risk of contamination or other process deviations [23].

There are various methods used for process monitoring of biopharmaceutical yeast fermentations, including online and offline techniques. Online methods involve the real-time measurement of key process variables using (e.g., optical) sensors, while offline methods involve the collection of samples for analysis in a laboratory setting [24].

Commonly used online monitoring techniques for biopharmaceutical yeast fermentations include: (1) in situ sensors for pH, dissolved oxygen, and temperature, (2) sensors for biomass and

ethanol, or (3) vibrational spectroscopy (e.g., NIR or IR) for in-line monitoring of multiple process variables [25–27]. While offline monitoring techniques may include: (1) high-performance liquid chromatography (HPLC) for quantification of metabolites and product concentration, (2) gas chromatography (GC) for monitoring of volatile metabolites, or (3) microscopic analysis for monitoring of cell morphology and viability. A significant amount of research in recent years has focused on developing novel process monitoring techniques for biopharmaceutical yeast fermentations, particularly those that can be used in real-time and provide more comprehensive information about the fermentation process, such as mass spectrometry and fluorescence spectroscopy. An example is the use of 2D EEM fluorescence, which has been used to successfully predict cell count and optical density (OD) in fed-batch *S. cerevisiae* fermentations [28].

These measurements, along with typical HPLC/UPLC, could be potentially used for model development, process monitoring, and control [23, 29, 30]. Two types of models are typically used in the industry, the mechanistic and the data driven. Mechanistic models require longer development time; however, they can be used from early development to manufacturing [31, 32]. The main use of mechanistic models is during process development for optimization, robustness, scale-up studies, and control analysis [33]. Furthermore, the use of mechanistic models for online monitoring has gained increasing attention recently [34]. On the other hand, data driven models are easily developed but they cannot be used extensively for scale-up studies and technology transfer to manufacturing facilities due the higher dependence on matrix-specific interactions. In a typical yeast fermentation process, samples are collected at most every few hours and usually not all the metabolites are monitored. Thus, use of an online instrument such as Raman, could provide very useful information for scientists and operators in near-real time.

1.1.1 | Raman Spectroscopy in Yeast Applications

Raman spectroscopy has significantly advanced the study of yeast fermentations, particularly those involving *S. cerevisiae*. While initial studies focused on ethanol fermentations, the use of Raman and chemometrics for on-line monitoring has expanded to include high-pressure fermentations, statistical process control, and even biorefinery applications. The first study to use Raman and chemometrics for on-line monitoring of ethanol production in *S. cerevisiae* fermentation was conducted by Shaw et al. [4]. Since then, this work has been built upon by many groups, including those by Picard et al. [35], who applied Raman to high-pressure fermentations, and Ávila et al. [36], who implemented statistical process control in glucose fermentation using Raman spectroscopy.

Moreover, Raman spectroscopy has been used to develop models for corn mash fermentations with glucose, oligosaccharides, and polysaccharides present [37], as well as in biorefinery (i.e., lignocellulosic) applications [7, 38, 39]. In more recent years, novel non-contact probes and feedback control schemes have been developed, providing researchers with even more tools to explore the complexities of *S. cerevisiae* fermentations. Notable examples include the development of a non-contact probe for in situ Raman measurements [40, 41], and the establishment of a

feedback control scheme to optimize *S. cerevisiae* fermentations for bioethanol production [42].

1.1.2 | Raman Spectroscopy in Biopharmaceutical Applications

Within yeast-based biopharmaceutical applications, a few studies using Raman spectroscopy exist. One study describes the development of calibration models on the primary substrates (glycerol and methanol) in supernatant samples from *Pichia pastoris* fermentations [43]. Here, the authors were not able to predict protein concentrations below 1 g/L. Others have investigated the use of Raman-enabled machine learning models as a tool for process control in *P. pastoris*-based fermentations for the production of a potential Malaria vaccine [44]. The models were calibrated using offline measurements on one Raman system, while the in situ measurements were obtained on another system. Moreover, the authors found that, due to non-linearities present within the data, support vector machines (SVM) outperformed partial least squares (PLS)-regression. This is common in systems with complex or highly fluorescent sample matrices [45, 46]. To our knowledge, there is no published literature showcasing the feasibility of using Raman for monitoring *S. cerevisiae* fermentations specifically in biopharmaceutical applications.

In the subsequent sections, a methodology is proposed for multivariate calibration of Raman spectra derived from cell-free media samples of *S. cerevisiae* fermentations. The models, calibrated using this approach, are evaluated based on prediction error to examine their ability to generalize.

2 | Materials and Methods

2.1 | Cultivation Data

2.1.1 | Yeast Strain and Medium

A recombinant *S. cerevisiae* strain is used in prolonged continuous cultivations. The recombinant strain expresses a heterologous protein. The heterologous protein is expressed extracellularly using the glycolytic triose-phosphate isomerase (TPI) promoter. The TPI promoter was chosen for the expression of the protein since it is considered strong [19]. It has furthermore been used with success in other studies for expression of heterologous proteins during prolonged continuous cultivations [47]. The cultivation medium was derived from ref. [48] with a supplement of yeast extract. No further details regarding the expression system, the heterologous protein, and the cultivation medium can be given due to confidentiality reasons.

2.1.2 | Inoculum

Inocula were prepared by transferring 450 μL cell culture to a 300 mL shake flask with 100 mL defined standard cultivation medium. The shake flasks were grown in a rotary shaker at 175 rpm overnight. The bioreactors were inoculated with the amount of seed culture corresponding to a start OD600 in the bioreactor around 0.5 AU.

2.1.3 | Standard Cultivations

Aerobic continuous cultivations were performed in 2 L fully instrumented and automatically controlled BIOSTAT B plus fermenters (Sartorius Stedim Biotech S.A., Germany) at a dilution rate of 0.08/h. Aerobic conditions were obtained by a stirrer speed of 1200 rpm and an aeration rate of 2 VVM. The cultivations were initiated with a fed-batch phase to increase the amount of biomass. The working volume in the continuous phase was kept at 1.2 L by a peristaltic effluent pump coupled to a weight measurement of the tank. pH was controlled at a constant level of 5.7 by addition of 10% NH_3.

2.1.4 | Sampling

Samples were taken daily for determination of protein concentration, residual glucose concentration, and extracellular metabolite concentration (Glucose, Glycerol, Acetate, Pyruvate, Succinate, *Metabolite X*). Checking for contamination, offline pH control, and verifying the volume in each bioreactor was also done on daily basis. Each sample was aliquoted into two cryo-tubes: one for reference analyses and one which was immediately frozen at −80°C. After the finalization of the experiments, all frozen samples were shipped for Raman analysis, as detailed in Section 2.2.

2.1.5 | Extracellular Metabolite Determination

Around 10 mL sample was taken from the cultivation broth using syringes containing frozen steel spheres. The steel spheres were frozen at −20°C. The sample was filtered through a 0.45 μm Sterivex filter (Merck, Darmstadt, Germany). Around 5 mL of filtered sample was needed. Syringes with frozen steel spheres were used to quench cellular metabolism [49]. The extracellular metabolites were measured by HPLC. The substrates in the samples were separated using an Aminex HPX-87H (Bio-Rad, California, USA) ion exclusion column, whereas an RI-detector and a UV-detector were used to determine the concentrations of the different substrates. Five millimolar H_2SO_4 in Milli-Q water was used as a mobile phase and the flow was set to 0.6 mL/min.

2.1.6 | Analysis of Recombinant Protein

The extracellular recombinant protein concentration was determined using 1 mL of cultivation broth. The broth was mixed with 3 mL KH_2PO_4 buffer to increase the pH to 12, followed by centrifugation at 5000 rpm for 10 min. The supernatant was then filtered and mixed with an equal volume of K_2HPO_4 buffer to neutralize the pH. Protein concentration was measured via HPLC using a Waters Alliance 2695 HPLC system equipped with a UV detector (Waters, Massachusetts, USA).

Metabolite X is a side-product synthesized by *S. cerevisiae* via the degradation pathway of a non-polar amino acid. Its presence is monitored in this study by a VIS-based spectroscopic method since a correlation between its synthesis and a lower productivity of the process has been observed.
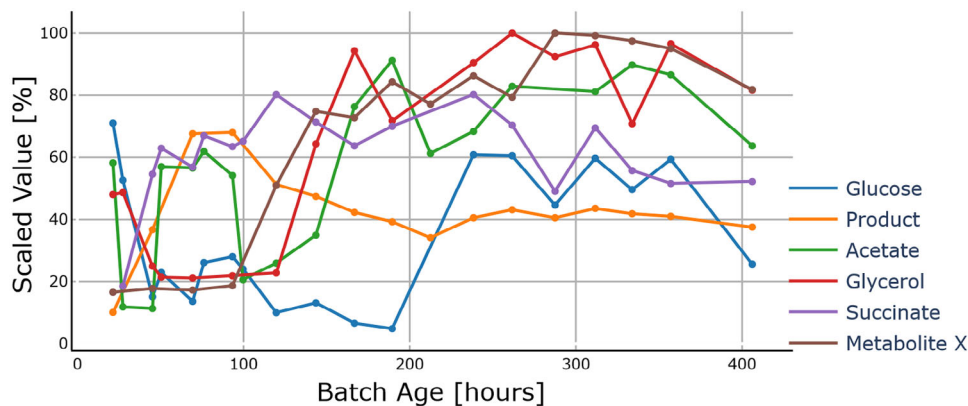
FIGURE 1 | Evolution of the six metabolites from Experiment 3 as a function of time. The *y*-axis is scaled to 100% of the overall values.

Figure 1 shows the evolution of metabolites as a function of fermentation time for Experiment 3, which is the longest running experiment. It was chosen to give an indication of how the metabolites vary over time. The values are scaled to the maximum of all the values. While not indicated in the figure, some of the metabolites have a few extreme samples, for example glucose, which were deemed out of scope for subsequent chemometric analysis, and thus removed prior to analysis.

2.2 | Raman Data Acquisition

The samples were delivered frozen and thawed at 5°C overnight. Before analysis, each sample tube was vortexed on the highest setting for 10 s each before being transferred to a quartz cuvette (P/N: CV10Q14, Thorlabs, Sweden). Raman spectra were collected using a Viserion 785 nm Analyzer (Indatech, Clapiers, France). The instrument uses a 180° backscatter configuration, and the detector was cooled to −60° to achieve optimal performance. Spectra were recorded using an integration time of 10 s, with 15 accumulations for a total acquisition time of 2.5 min/spectra. Automatic cosmic ray filtering was used, but otherwise the Raman spectra were not modified in any way before data analysis.

The experimental data encompass a range of fermentation durations and number of spectra collected. On average, each experiment included around 18 spectra, with run times varying from 144 to 406 h. This variation was not intentional but provides a broad dataset to validate the Raman spectroscopy method. For instance, Experiment 1 had 11 spectra over 165 h, while Experiment 3 spanned 406 h and included 34 spectra. This diverse dataset allows for a comprehensive analysis of the Raman spectroscopy's performance across different conditions.

2.3 | Chemometric Modelling

2.3.1 | Data Pretreatment

The pretreatment procedure is essential to improve the signal-to-noise ratio and comprises several important operations [50, 51]. In this analysis, motivated by the approach established in the commercially available software DataHowLab Spectra (DataHow, Zurich, Switzerland), the following pretreatment steps were used:

Variable Selection–Selection of the regions of the spectra that are most likely to contain relevant information [52]. As a baseline, the following Raman shift regions were used: 450–3050 cm^{-1}, with the region 1800–2800 cm^{-1} excluded.

Furthermore, variable selection can be performed algorithmically. Several algorithms exist to solve this without the need for a priori knowledge [53, 54]. In this study, we use the variable importance in projection (VIP)-algorithm as an example of variable selection [55] since it is fast and simple. However, many more exist. For a comprehensive view of available methods, the readers are encouraged to review ref. [52]. In general, these methods can be used to denoise spectral signals by explicitly selecting the most relevant wavenumbers for the prediction task.

Savitzky-Golay (SG)–A smoothing and differentiation technique that fits a moving polynomial to the spectra to reduce noise or irregularities [56]. This is done along the wavenumber dimension and has three integer parameters: (1) Window length: number of consecutive features (wavenumbers) to consider in the moving window (testing window sizes between 11 and 131 in steps of 10). (2) Polynomial order: order of the fitted polynomial (testing either second, third, or fourth order). (3) Derivative order: order of the signal to compute (integer values between zeroth and second order were considered).

Row-wise scaling–Using standard normal variate (SNV) to scale the rows of **X** to mean zero and a standard deviation of one [57]. This method works well when most of the constituent signal is the same in the spectra [58].

Column-wise centering and scaling. A set of simple methods that makes the model treat all wavenumbers similarly [59]. This preprocessing operation has three Boolean options: (1) Mean-centering: Remove the mean across samples for each wavenumber. (2) Scaling: Divide spectra values by the standard deviation across samples. (3) Run-wise: Compute mean and standard deviation with respect to each of the subsets.

The preprocessing approach used the following sequence: Savitsky-Golay smoothing/derivative, wavenumber selection, SNV followed by a centering and scaling step. The optimization of the hyperparameters was performed automatically during

modelling. The search ranges for the various preprocessing steps were as follows: For Savitsky-Golay; window size of 11–131, polynomial orders of two, three and four, derivative orders between zero and two. For the scaling step, we considered mean-centering, autoscaling, and run-wise scaling.

2.3.2 | Partial Least Squares Regression

The partial least squares (PLS) model is a multivariate bi-linear algorithm that projects the highly correlated input **X** and output **Y** matrices unto new uncorrelated latent spaces, the scores of the input and output matrices [60, 61]. Simultaneously it performs a regression between the scores to maximize the covariance between **X** and **Y** [62]. PLS is a widely utilized method in chemometrics for calibrating models based on spectra [63]. This work uses the PLS1 algorithm, where the independent variable, **y**, is univariate, that is, a vector. Thus, for each of the six variables of interest in this study, a single model will be calibrated to yield a total of six models. The number of PLS components considered in the study was between one and six.

To assess if further improvement to the prediction results was possible, VIP-scores were employed as a variable selection tool [64]. In this study, a threshold of 0.8 is used for VIP-based variable selection, meaning variables with VIP-scores above this threshold are considered unlikely to represent noise and thus retained in the model [65].

Additionally, attempts were made to test other machine learning methodologies, such as SVM and extra trees regressors (ETR). These methodologies did not outperform PLS-VIP (data not shown). It is likely due to either: (1) the datasets being too small to achieve adequate convergence [66], or (2) the data can be assumed to be linear and as such, a bilinear model like PLS will be able to explain the variation in the data sufficiently.

2.3.3 | Modelling Workflow

The proposed methodology utilizes a nested cross-validation (CV) approach to ascertain the model's ability to generalize across different data sets [67, 68]. In the outer loop of this approach, one experiment is isolated from the others to serve as a test run, thereby preventing information leakage between the sets. The inner loop fine-tunes five hyperparameters of the preprocessing pipeline (window length, polynomial order, derivative order, scaling, and the number of latent variables) using the training data. This fine-tuning process combines five-fold CV and a guided optimization loop.

Five-fold CV offers a reliable estimate of the model's performance on unseen data, whereas the guided optimization loop implements a strategic sampling strategy for the next set of hyperparameters [69]. By balancing exploration and exploitation of the hyperparameter space, this strategy aims for efficient convergence to an optimal solution, expressed as the minimal root mean squared error of cross-validation (RMSECV). After this nested optimization, the final hyperparameters chosen are those leading to the lowest RMSECV. These are then used to retrain

the model on the complete calibration data set. Within the inner CV loop, the incorporation of VIP selection (the PLS-VIP method) initiates by calculating the VIP values for all wavenumbers using the trained PLS model. Those with a VIP value exceeding 0.8 are then identified. Subsequently, a second optimization campaign is initiated, aimed at determining the optimal hyperparameters for this feature subset. The selected hyperparameters are then used to retrain the model on the training set and evaluate it on the test set, but only incorporating the wavenumbers identified during the VIP selection process. In cases where VIP selection is not applied, the model is directly evaluated on the test set after the initial optimization campaign, making use of all input features. In this study, a comparison is made between the performance of the PLS method with and without the use of VIP selection.

Given there was only data available for six experiments, the training process for both methods was repeated by iterating the outer CV loop with a new experiment serving as the test run, and the respective training procedure was repeated on a new training set. This strategy facilitated the evaluation of the proposed approach's sensitivity to a limited number of calibration data points. It also examined its generalization capabilities when applied to disparate test experiments, by calculating an average prediction error across all possible runs when each was used individually as a test set.

A visual representation of the modeling workflow employed in this study is presented in Figure 2. It consists of the following steps: **Train/test split**, where the six experiments are iteratively split into five train runs and one test run. **Optimization**, where the preprocessing and other hyperparameters are evaluated to converge on the best solution under five-fold CV (1 experiment of 5 held out each time). **Retraining**, where the optimal parameters found previously are applied to the entire train set. **Evaluation**, where finally, the trained model is applied to predict on the unseen test run. The process depicted was repeated six times, to ensure each experiment is used once in model evaluation. In the cases where **VIP-selection** was used, the modeling workflow consisted of additional steps to compute VIP-scores before Optimization and Retraining with just the selected wavenumbers.

2.3.4 | Evaluation of Model Performance

The performance of the models is evaluated using the root mean squared error (RMSE)-statistic. It is preferred over the coefficient of determination (R^2) due to the original units being retained and thus making it easier for interpretation. The root mean squared error of prediction (RMSEP) is used for evaluation.

The RMSEP is calculated by the following equation (Equation 1):

$$\text{Relative RMSEP} = \frac{\sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}}{s(y)} \qquad (1)$$

In Equation (1), n represents the number of samples, y_i and $y(hat)_i$ denote the measured and model-predicted offline concentrations for the i-th sample, respectively, and $s(y)$ is the standard deviation
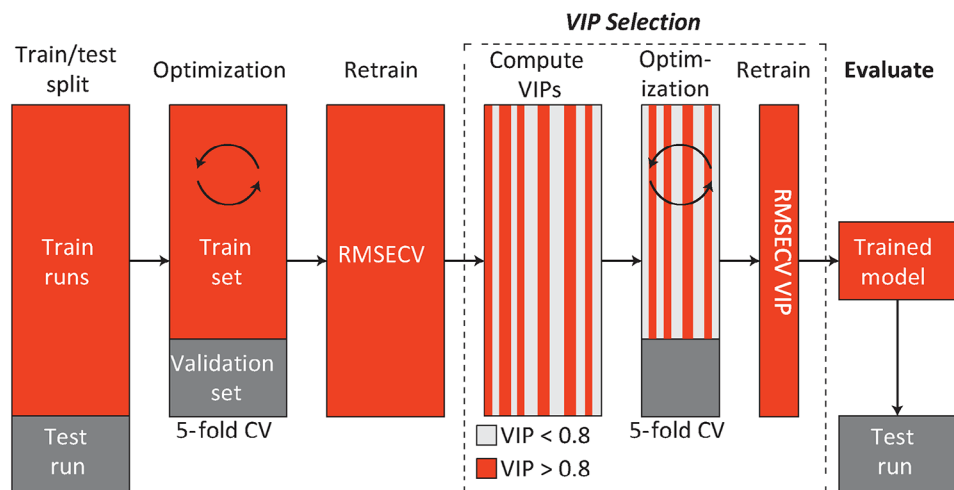
FIGURE 2 | Modeling workflow used to optimize hyperparameters, calibrate, and evaluate on the test run. The dashed area, representing the VIP selection block, was only used for PLS-VIP model.

of the variable y. The final step involves computing the average of the errors from the six test sets, which is then utilized for comparison. This number represents the average prediction error for an unseen experiment, scaled by the standard deviation of the respective variable.

3 | Results and Discussion

The PLS-VIP model consistently outperformed the standard PLS across all process variables, as evidenced by the RMSEP. Notably, the PLS-VIP model achieved the lowest VIP selection average relative RMSEP errors for Product and *Metabolite X*, both below 0.4 g/L. Conversely, Glucose exhibited the largest average error at 0.82 g/L, likely due to the limited range of observed glucose values and the reduced sensitivity of Raman spectroscopy at these concentrations.

For example, the RMSEP for Glucose was reduced from 0.90 g/L (SD 0.32) to 0.82 g/L (SD 0.25) when using the PLS-VIP model. Similarly, for Product, the error decreased from 0.38 g/L (SD 0.09) to 0.35 g/L (SD 0.08). Acetate also saw an improvement, with errors reducing from 0.88 g/L (SD 0.21) to 0.77 g/L (SD 0.16). Glycerol and Succinate showed smaller but still notable improvements. Specifically, Glycerol's RMSEP reduced from 0.62 g/L (SD 0.24) to 0.53 g/L (SD 0.20), and Succinate's RMSEP slightly improved from 0.73 g/L (SD 0.24) to 0.72 g/L (SD 0.23). For *Metabolite X*, the RMSEP decreased from 0.40 g/L (SD 0.15) to 0.36 g/L (SD 0.13).

The standard deviation of the prediction errors indicates the robustness of the models. Both Product and *Metabolite X* exhibited low standard deviations, suggesting that the proposed methodology is both accurate and robust to changes in calibration and test data. The PLS-VIP model consistently demonstrated marginally better robustness than the standard PLS, likely due to its parsimonious nature.

The results indicate that the general trend for both Product and *Metabolite X* (Figure 3B,F) are well captured by the cali-

bration model for all experiments. For Glycerol and Succinate (Figure 3D,E), the deviation from the optimal line can vary among different test experiments. This observation can be attributed to the limitations of data-driven models in extrapolating beyond the calibration range. When the experimental values of the test set fall outside the range observed during calibration, the prediction errors tend to increase. The same can be observed when predicting *Metabolite X* for Experiment 3, where the model mostly under-predicts when extrapolating, as seen in Figure 3F, where most of the points above 30% are below the 1:1 line. The extent of this variability is expected to diminish as larger training sets are utilized following an increasing number of fermentations monitored by Raman, covering a more comprehensive range of process behavior. Moreover, the vertical line of points seen at the lower end of the observed scale for *Metabolite X* and Glycerol might be an indication of the lower detection limit of the calibration model.

The process evolution plots depicted in Figure 4 validate the usefulness of such soft sensors in monitoring and control applications. This is particularly evident in cases where there are rapid changes in the process dynamics, such as sudden accumulation or consumption of a metabolite, and the model consistently tracks these changes over time. Moreover, for Experiments 3 and 4 (Figure 4B,E), the models show good predictive performance for both variables up to 400 h of process run time, where the spectra acquired are affected by high fluorescence effects.

While the generated models perform well in general, Figure 4 shows that, at low concentrations, model performance is poor for Experiment 6 (Figure 4C,F). This could be due to several factors, including the inherent sensitivity limits of the Raman technique and the reduced signal-to-noise ratio at low concentrations. Further studies are needed to refine the model and improve its performance at low concentrations.

These results suggest that it might be possible to predict process variables of interest based on Raman spectral information collected in a bioreactor fermentation process using the proposed methodology. In other words, based on the utilization of a single
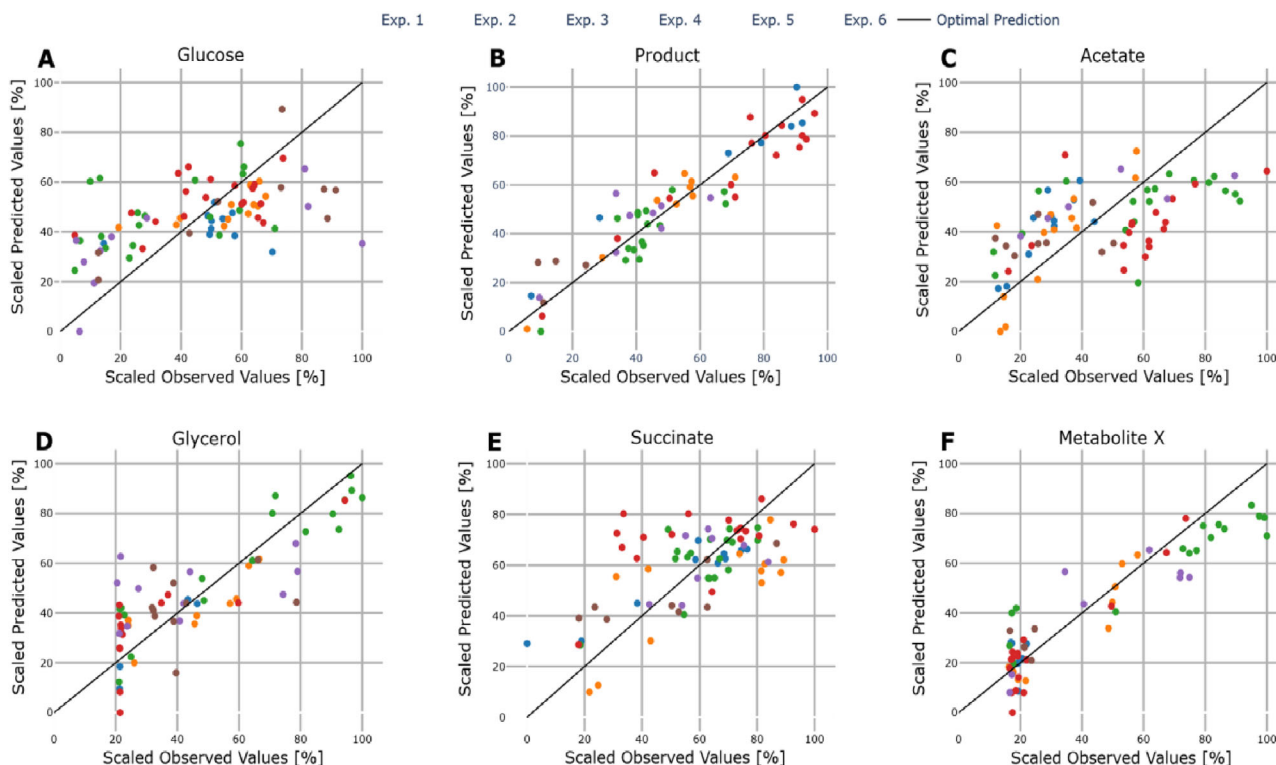
FIGURE 3 | Observed versus predicted plots of PLS-VIP performance for each target variable on all experiments (when in the test set/held out).
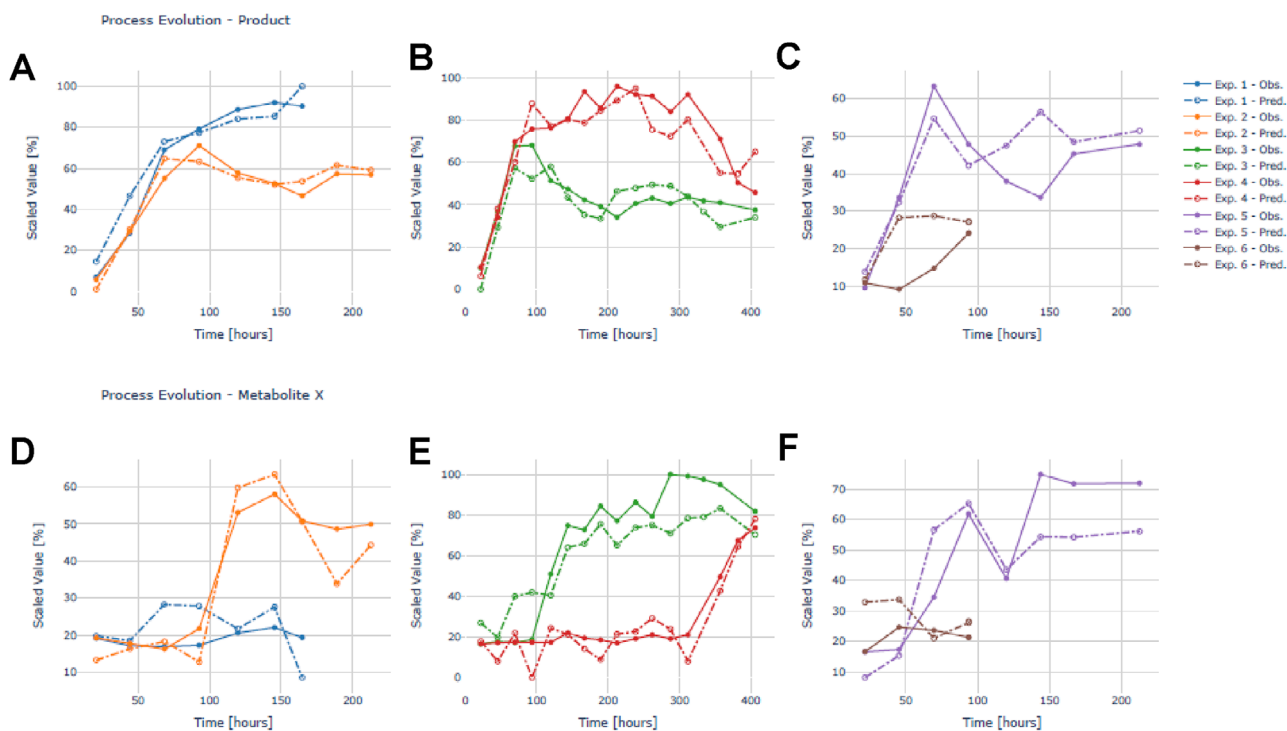

FIGURE 4 | Process evolution plots for Product (A–C) and *Metabolite X* (D–F) where observed values correspond to the real samples and predicted values (dashed lines) correspond to predictions using PLS-VIP model.

sensor, the proposed methodology could allow for the quantification of numerous variables such as the Product, *Metabolite X*, and Glycerol accurately. This is achieved by the effective elimination of the strong fluorescent background by filtering the cells and by data preprocessing. By employing VIP selection, the spectra can

be further denoised, enhancing the learning ability and predictive performance of the model.

The work of others (e.g., Bogomolov et al. [28]) underscores the fact that relevant chemical information can be found by

spectroscopic techniques in general. In this instance, the authors show that the strong fluorescence observed in yeast-based matrices lends itself well to using 2D fluorescence spectra. This underlines the hypothesis that removing the cells prior to Raman analysis will lead to better results.

Besides fluorescence spectroscopy, NIR has been shown to produce good results in in situ applications of ethanol fermentations, even for some of the same metabolites as those investigated presently [27]. The reported standard errors of prediction were 0.51 and 0.79 g/L for glycerol and glucose, respectively—two metabolites also investigated in this analysis. The values are quite comparable, with RMSEP of 0.53 and 0.82 g/L using Raman spectroscopy. It is currently unknown whether the use of NIR will translate to the more complex biopharmaceutical fermentations. However, internal investigations have indicated that probe fouling remains a significant issue (data not shown).

Since all six fermentations were designed with similar conditions, there are no guarantees at this stage that the proposed approach is not also correcting a random quantity exhibiting a behavior like the targets. Ongoing work is focused on breaking such correlations via a spiking study, as has been explored in the literature [70].

One of the key advantages of Raman spectroscopy is its potential for in line or in situ deployment, allowing real-time monitoring within the fermenter. In a chemostat, a continuous fermentation system utilized in this study, the harvest flow can be continuously filtered to remove cells. This setup enables effective at-line monitoring, providing real-time data on metabolite concentrations without autofluorescence interference. By monitoring the fermentation from the harvest flow, we can achieve precise control and optimization of the process. Moreover, some components measured here, and others not included in this work, are currently analyzed by more laborious methods than HPLC, limiting data acquisition from each fermentation. Additionally, some analytes require even more cumbersome methods than HPLC, further highlighting the advantages of Raman spectroscopy. Moreover, the potential for Raman to replace the entire suite of reference instruments, providing a single point-of-analysis, remains attractive regardless of the interfacing method. If in line deployment is not achieved with continuous filtration (such as that used in perfusion in mammalian cells), an at-line Raman system can still service multiple bioreactors without suffering from calibration transfer issues between probes or instruments.

4 | Conclusions

This study demonstrates the potential for accurately predicting critical process quantities using Raman spectroscopy in cell-free media samples from *S. cerevisiae* fermentations. Future studies should consider spiking samples with specific quantities of metabolites to better assess the method's accuracy, uniqueness, and linearity. The findings indicate that Raman spectroscopy could play an increasingly significant role in monitoring and controlling microbial fermentations, leading to enhanced process efficiency and product quality.

5 | Future Work

For future work, it could be advantageous to explore the integration of the proposed chemometric methodology with hybrid models that predict process dynamics using a filtering approach [71]. Incorporating ensemble techniques like bagging could prove beneficial when assessing model uncertainty in prediction. Additionally, a spiking design could enhance model training when only a few training runs are available, especially in regions where available samples are limited. This design could address the challenges posed by scarce data and improve model performance in such scenarios.

On the process side, working to either integrate the Raman probe directly into the process, for example, by innovative probe design or by a continuous filtering setup where the Raman signal is collected from the permeate, could also prove very beneficial. Alternate tangential filters (ATFs) are now commonplace in mammalian culturing, and the advent of these in biopharmaceutical fermentations could unlock new possibilities for PAT.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

References

1. A. S. Rathore, R. Bhambure, and V. Ghare, "Process Analytical Technology (PAT) for Biopharmaceutical Products," *Analytical and Bioanalytical Chemistry* 398 (2010): 137–154, https://doi.org/10.1007/s00216-010-3781-x.

2. J. Rantanen, "Process Analytical Applications of Raman Spectroscopy," *Journal of Pharmacy and Pharmacology* 59 (2010): 171–177, https://doi.org/10.1211/jpp.59.2.0004.

3. K. A. Bakeev, *Process Analytical Technology* 2nd ed, (Chichester, UK: Wiley, 2010).

4. A. D. Shaw, N. Kaderbhai, A. Jones, et al., "Noninvasive, on-Line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol

Using Dispersive Raman Spectroscopy and Chemometrics," *Applied Spectroscopy* 53 (1999): 1419–1428, https://doi.org/10.1366/0003702991945777.

5. A. C. McGovern, D. Broadhurst, J. Taylor, et al., "Monitoring of Complex Industrial Bioprocesses for Metabolite Concentrations Using Modern Spectroscopies and Machine Learning: Application to Gibberellic Acid Production," *Biotechnology and Bioengineering* 78 (2002): 527–538, https://doi.org/10.1002/bit.10226.

6. H. L. Lee, P. Boccazzi, N. Gorret, R. J. Ram, and A. J. Sinskey, "In Situ Bioprocess Monitoring of *Escherichia coli* Bioreactions Using Raman Spectroscopy," *Vibrational Spectroscopy* 35 (2004): 131–137, https://doi.org/10.1016/j.vibspec.2003.12.015.

7. S. Ewanick, E. Schmitt, R. Gustafson, and R. Bura, "Use of Raman Spectroscopy for Continuous Monitoring and Control of Lignocellulosic Biorefinery Processes," *Pure and Applied Chemistry* 86 (2014): 867–879, https://doi.org/10.1515/pac-2013-1022.

8. B. Li, P. W. Ryan, B. H. Ray, K. J. Leister, N. M. Sirimuthu, and A. G. Ryder, "Rapid Characterization and Quality Control of Complex Cell Culture media Solutions Using Raman Spectroscopy and Chemometrics," *Biotechnology and Bioengineering* 107 (2010): 290–301, https://doi.org/10.1002/bit.22813.

9. N. R. Abu-Absi, B. M. Kenty, M. E. Cuellar, et al., "Real Time Monitoring of Multiple Parameters in Mammalian Cell Culture Bioreactors Using an In-Line Raman Spectroscopy Probe," *Biotechnology and Bioengineering* 108 (2011): 1215–1221, https://doi.org/10.1002/bit.23023.

10. K. Buckley and A. G. Ryder, "Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review," *Applied Spectroscopy* 71 (2017): 1085–1116, https://doi.org/10.1177/0003702817703270.

11. K. A. Esmonde-White, M. Cuellar, and I. R. Lewis, "The Role of Raman Spectroscopy in Biopharmaceuticals From Development to Manufacturing," *Analytical and Bioanalytical Chemistry* 414 (2022): 969–991, https://doi.org/10.1007/s00216-021-03727-4.

12. J. B. Cooper, "Chemometric Analysis of Raman Spectroscopic Data for Process Control Applications," *Chemometrics and Intelligent Laboratory Systems* 46 (1999): 231–247, https://doi.org/10.1016/S0169-7439(98)00174-9.

13. O. Ryabchykov, S. Guo, and T. Bocklitz, "Analyzing Raman Spectroscopic Data," *Physical Sciences Reviews* 4 (2019): 1–16, https://doi.org/10.1515/psr-2017-0043.

14. J. M. Bielser, M. Wolf, J. Souquet, H. Broly, and M. Morbidelli, "Perfusion Mammalian Cell Culture for Recombinant Protein Manufacturing—A Critical Review," *Biotechnology Advances* 36 (2018): 1328–1340, https://doi.org/10.1016/j.biotechadv.2018.04.011.

15. Z. Liu, Z. Zhang, Y. Qin, et al., "The Application of Raman Spectroscopy for Monitoring Product Quality Attributes in Perfusion Cell Culture," *Biochemical Engineering Journal* 173 (2021): 108064, https://doi.org/10.1016/j.bej.2021.108064.

16. D. Széliová, D. E. Ruckerbauer, S. N. Galleguillos, et al., "What CHO Is Made of: Variations in the Biomass Composition of Chinese Hamster Ovary Cell Lines," *Metabolic Engineering* 61 (2020): 288–300, https://doi.org/10.1016/j.ymben.2020.06.002.

17. K. K. Frame and W. Hu, "Cell Volume Measurement as an Estimation of Mammalian Cell Biomass," *Biotechnology and Bioengineering* 36 (1990): 191–197, https://doi.org/10.1002/bit.260360211.

18. T. E. Matthews, J. P. Smelko, B. Berry, et al., "Glucose Monitoring and Adaptive Feeding of Mammalian Cell Culture in the Presence of Strong Autofluorescence by Near Infrared Raman Spectroscopy," *Biotechnology Progress* 34 (2018): 1574–1580, https://doi.org/10.1002/btpr.2711.

19. Z. Liu, K. E. Tyo, J. L. Martínez, D. Petranovic, and J. Nielsen, "Different Expression Systems for Production of Recombinant Proteins in Saccharomyces Cerevisiae," *Biotechnology and Bioengineering* 109 (2012): 1259–1268, https://doi.org/10.1002/bit.24409.

20. A. G. Ryder, "Cell Culture Media Analysis Using Rapid Spectroscopic Methods," *Current Opinion in Chemical Engineering* 22 (2018): 11–17, https://doi.org/10.1016/j.coche.2018.08.008.

21. E. Wells and A. S. Robinson, "Cellular Engineering for Therapeutic Protein Production: Product Quality, Host Modification, and Process Improvement," *Biotechnology Journal* 12 (2017): 1600105, https://doi.org/10.1002/biot.201600105.

22. J. Surre, C. Saint-Ruf, V. Collin, S. Orenga, M. Ramjeet, and I. Matic, "Strong Increase in the Autofluorescence of Cells Signals Struggle for Survival," *Scientific Reports* 8 (2018): 12088, https://doi.org/10.1038/s41598-018-30623-2.

23. L. Mears, S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey, "Mechanistic Fermentation Models for Process Design, Monitoring, and Control," *Trends in Biotechnology* 35 (2017): 914–924, https://doi.org/10.1016/j.tibtech.2017.07.002.

24. R. Ulber, J. G. Frerichs, and S. Beutel, "Optical Sensor Systems for Bioprocess Monitoring," *Analytical and Bioanalytical Chemistry* 376 (2003): 342–348, https://doi.org/10.1007/s00216-003-1930-1.

25. J. Randek, *Advancement of Sensor Technology for Monitoring and Control of Upstream Bioprocesses. 2056* (Linköping, Sweden: Linköping University Electronic Press, 2020).

26. K. Pontius, H. Junicke, K. V. Gernaey, and M. Bevilacqua, "Monitoring Yeast Fermentations by Nonlinear Infrared Technology and Chemometrics—Understanding Process Correlations and Indirect Predictions," *Applied Microbiology and Biotechnology* 104 (2020): 5315–5335, https://doi.org/10.1007/s00253-020-10604-0.

27. V. A. Corro-Herrera, J. Gómez-Rodríguez, P. M. Hayward-Jones, D. M. Barradas-Dermitz, M. G. Aguilar-Uscanga, and A. C. Gschaedler-Mathis, "In-Situ Monitoring of Saccharomyces Cerevisiae ITV01 Bioethanol Process Using Near-Infrared Spectroscopy NIRS and Chemometrics," *Biotechnology Progress* 32 (2016): 510–517, https://doi.org/10.1002/btpr.2222.

28. A. Bogomolov, T. Grasser, and M. Hessling, "In-Line Monitoring of Saccharomyces Cerevisiae Fermentation With a Fluorescence Probe: New Approaches to Data Collection and Analysis," *Journal of Chemometrics* 25 (2011): 389–399, https://doi.org/10.1002/cem.1365.

29. C. F. Mandenius, "Recent Developments in the Monitoring, Modeling and Control of Biological Production Systems," *Bioprocess and Biosystems Engineering* 26 (2004): 347–351, https://doi.org/10.1007/s00449-004-0383-z.

30. M. M. Schuler and I. W. Marison, "Real-Time Monitoring and Control of Microbial Bioprocesses With Focus on the Specific Growth Rate: Current State and Perspectives," *Applied Microbiology and Biotechnology* 94 (2012): 1469–1482, https://doi.org/10.1007/s00253-012-4095-z.

31. D. Carranza-Saavedra, C. P. S. Henao, and J. E. Z. Montoya, "Kinetic Analysis and Modeling of L-Valine Production in Fermentation Batch From *E. coli* Using Glucose, Lactose and Whey as Carbon Sources," *Biotechnology Reports* 31 (2021): e00642, https://doi.org/10.1016/j.btre.2021.e00642.

32. A. B. R. de Jesus and J. C. de Carvalho Miranda, "Assessment of the Applicability of Methods and Tools in Process Systems Engineering for Fermentation Processes," *Biofuels, Bioproducts and Biorefining* 17 (2023): 1121–1155, https://doi.org/10.1002/bbb.2520.

33. S. Espinel-Ríos, K. Bettenbrock, S. Klamt, and R. Findeisen, "Maximizing Batch Fermentation Efficiency by Constrained Model-Based Optimization and Predictive Control of Adenosine Triphosphate Turnover," *AIChE Journal* 68 (2022), https://doi.org/10.1002/aic.17555.

34. P. Shah, M. Z. Sheriff, M. S. F. Bangi, et al., "Deep Neural Network-Based Hybrid Modeling and Experimental Validation for an Industry-Scale Fermentation Process: Identification of Time-Varying Dependencies Among Parameters," *Chemical Engineering Journal* 441 (2022): 135643, https://doi.org/10.1016/j.cej.2022.135643.

35. A. Picard, I. Daniel, G. Montagnac, and P. Oger, "In Situ Monitoring by Quantitative Raman Spectroscopy of Alcoholic Fermentation by Saccharomyces Cerevisiae Under High Pressure," *Extremophiles* 11 (2007): 445–452, https://doi.org/10.1007/s00792-006-0054-x.

36. T. C. Ávila, R. J. Poppi, I. Lunardi, P. A. G. Tizei, and G. A. G. Pereira, "Raman Spectroscopy and Chemometrics for On-Line Control of Glucose Fermentation by Saccharomyces Cerevisiae," *Biotechnology Progress* 28 (2012): 1598–1604, https://doi.org/10.1002/btpr.1615.

37. S. R. Gray, S. W. Peretti, and H. H. Lamb, "Real-Time Monitoring of High-Gravity Corn Mash Fermentation Using In-Situ Raman Spectroscopy," *Biotechnology and Bioengineering* 110 (2013): 1654–1662, https://doi.org/10.1002/bit.24849.

38. J. A. Iversen, R. W. Berg, and B. K. Ahring, "Quantitative Monitoring of Yeast Fermentation Using Raman Spectroscopy," *Analytical and Bioanalytical Chemistry* 406 (2014): 4911–4919, https://doi.org/10.1007/s00216-014-7897-2.

39. J. A. Iversen and B. K. Ahring, "Monitoring Lignocellulosic Bioethanol Production Processes Using Raman Spectroscopy," *Bioresource Technology* 172 (2014): 112–120, https://doi.org/10.1016/j.biortech.2014.08.068.

40. R. Schalk, F. Braun, R. Frank, et al., "Non-Contact Raman Spectroscopy for In-Line Monitoring of Glucose and Ethanol During Yeast Fermentations," *Bioprocess and Biosystems Engineering* 40 (2017): 1519–1527, https://doi.org/10.1007/s00449-017-1808-9.

41. R. Schalk, A. Heintz, F. Braun, et al., "Comparison of Raman and Mid-Infrared Spectroscopy for Real-Time Monitoring of Yeast Fermentations: A Proof-of-Concept for Multi-Channel Photometric Sensors," *Applied Sciences* 9 (2019): 2472, https://doi.org/10.3390/app9122472.

42. E. Hirsch, H. Pataki, J. Domján, et al., "Inline Noninvasive Raman Monitoring and Feedback Control of Glucose Concentration During Ethanol Fermentation," *Biotechnology Progress* 35 (2019): 1–8, https://doi.org/10.1002/btpr.2848.

43. A. Paul, P. Carl, F. Westad, J. Voss, and M. Maiwald, "Towards Process Spectroscopy in Complex Fermentation Samples and Mixtures," *Chemie Ingenieur Technik* 88 (2016): 756–763, https://doi.org/10.1002/cite.201500118.

44. J. Voss, N. E. Mittelheuser, R. Lemke, and R. Luttmann, "Advanced Monitoring and Control of Pharmaceutical Production Processes With *Pichia pastoris* by Using Raman Spectroscopy and Multivariate Calibration Methods," *Engineering in Life Sciences* 17 (2017): 1281–1294, https://doi.org/10.1002/elsc.201600229.

45. J. Hagedorn, C. Halloin, E. Skibsted, L. Poulsen, and M. A. Hedegaard, "Determination of Stem Cell Pluripotency in Spent Cultivation Media by Raman Spectroscopy," *Journal of Raman Spectroscopy* (2023): 608–618, https://doi.org/10.1002/jrs.6532.

46. C. Rafferty, K. Johnson, J. O'Mahony, B. Burgoyne, R. Rea, and K. M. Balss, "Analysis of Chemometric Models Applied to Raman Spectroscopy for Monitoring Key Metabolites of Cell Culture," *Biotechnology Progress* 36 (2020): 1–16, https://doi.org/10.1002/btpr.2977.

47. A. K. Seresht, A. L. Cruz, d. E. Hulster, et al., "Long-Term Adaptation of Saccharomyces Cerevisiae to the Burden of Recombinant Insulin Production," *Biotechnology and Bioengineering* 110 (2013): 2749–2763, https://doi.org/10.1002/bit.24927.

48. C. Verduyn, "Physiology of Yeasts in Relation to Biomass Yields," *Antonie Van Leeuwenhoek* 60 (1991): 325–353, https://doi.org/10.1007/BF00430373.

49. M. Mashego, M. Jansen, J. Vinke, W. Vangulik, and J. Heijnen, "Changes in the Metabolome of Associated With Evolution in Aerobic Glucose-Limited Chemostats," *FEMS Yeast Research* 5 (2005): 419–430, https://doi.org/10.1016/j.femsyr.2004.11.008.

50. J. M. Roger, J. C. Boulet, M. Zeaiter, and D. N Rutledge, "Pre-Processing Methods," in *Comprehensive Chemometrics*, 2nd edn. eds: S. Brown, R. Tauler and B. Walczak (Amsterdam, Netherlands: Elsevier, 2020), pp. 1–75), https://doi.org/10.1016/B978-0-12-409547-2.14878-4.

51. S. Guo, J. Popp, and T. Bocklitz, "Chemometric Analysis in Raman Spectroscopy From Experimental Design to Machine Learning–Based Modeling," *Nature Protocols* 16 (2021): 5426–5459, https://doi.org/10.1038/s41596-021-00620-3.

52. T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A Review of Variable Selection Methods in Partial Least Squares Regression," *Chemometrics and Intelligent Laboratory Systems* 118 (2012): 62–69, https://doi.org/10.1016/j.chemolab.2012.07.010.

53. T. Mehmood, S. Sæbø, and K. H. Liland, "Comparison of Variable Selection Methods in Partial Least Squares Regression," *Journal of Chemometrics* 34 (2020), https://doi.org/10.1002/cem.3226.

54. H. Jiang, W. Xu, Y. Ding, and Q. Chen, "Quantitative Analysis of Yeast Fermentation Process Using Raman Spectroscopy: Comparison of CARS and VCPA for Variable Selection," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 228 (2020): 117781, https://doi.org/10.1016/j.saa.2019.117781.

55. I. G. Chong and C. H. Jun, "Performance of Some Variable Selection Methods When Multicollinearity is Present," *Chemometrics and Intelligent Laboratory Systems* 78 (2005): 103–112, https://doi.org/10.1016/j.chemolab.2004.12.011.

56. A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry* 36 (1964): 1627–1639, https://doi.org/10.1021/ac60214a047.

57. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra," *Applied Spectroscopy* 43 (1989): 772–777, https://doi.org/10.1366/0003702894202201.

58. M. Dhanoa, S. Lister, R. Sanderson, and R. Barnes, "The Link Between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra," *Journal of Near Infrared Spectroscopy* 2 (1994): 43–47, https://doi.org/10.1255/jnirs.30.

59. A. Smolinska, L. Blanchet, L. M. Buydens, and S. S. Wijmenga, "NMR and Pattern Recognition Methods in Metabolomics: From Data Acquisition to Biomarker Discovery: A Review," *Analytica Chimica Acta* 750 (2012): 82–97, https://doi.org/10.1016/j.aca.2012.05.049.

60. S. Wold, M. Sjöström, and L. Eriksson, "PLS-Regression: A Basic Tool of Chemometrics," *Chemometrics and Intelligent Laboratory Systems* 58 (2001): 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.

61. A. Höskuldsson, "PLS Regression Methods," *Journal of Chemometrics* 2 (1988): 211–228, https://doi.org/10.1002/cem.1180020306.

62. P. Geladi and B. R. Kowalski, "Partial Least-Squares Regression: A Tutorial," *Analytica Chimica Acta* 185 (1986): 1–17, https://doi.org/10.1016/0003-2670(86)80028-9.

63. H. Abdi, "Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression)," *WIREs Computational Statistics* 2 (2010): 97–106, https://doi.org/10.1002/wics.51.

64. C. M. Andersen and R. Bro, "Variable Selection in Regression—A Tutorial," *Journal of Chemometrics* 24 (2010): 728–737, https://doi.org/10.1002/cem.1360.

65. N. Akarachantachote, S. Chadcham, and K. Saithanu, "Cutoff Threshold of Variable Importance in Projection for Variable Selection," *International Journal of Pure and Applied Mathematics* 94 (2014): 307–322, 10.12732/ij-pam.v94i3.2.

66. R. I. Tange, M. A. Rasmussen, E. Taira, and R. Bro, "Benchmarking Support Vector Regression Against Partial Least Squares Regression and Artificial Neural Network: Effect of Sample Size on Model Performance," *Journal of Near Infrared Spectroscopy* 25 (2017): 381–390, https://doi.org/10.1177/0967033517734945.

67. F. Westad and F. Marini, "Validation of Chemometric Models–A Tutorial," *Analytica Chimica Acta* 893 (2015): 14–24, https://doi.org/10.1016/j.aca.2015.06.056.

68. E. Anderssen, K. Dyrstad, F. Westad, and H. Martens, "Reducing Over-Optimism in Variable Selection by Cross-Model Validation," *Chemometrics and Intelligent Laboratory Systems* 84 (2006): 69–74, https://doi.org/10.1016/j.chemolab.2006.04.021.

69. S. Guo, T. Bocklitz, U. Neugebauer, and J. Popp, "Common Mistakes in Cross-Validating Classification Models," *Analytical Methods* 9 (2017): 4410–4417, https://doi.org/10.1039/C7AY01363A.

70. P. Romann, J. Kolar, D. Tobler, C. Herwig, J. Bielser, and T. K. Villiger, "Advancing Raman Model Calibration for Perfusion Bioprocesses Using Spiked Harvest Libraries," *Biotechnology Journal* 17 (2022): 1–14, https://doi.org/10.1002/biot.202200184.

71. H. Narayanan, M. F. Luna, M. von Stosch, et al., "Bioprocessing in the Digital Age: The Role of Process Models," *Biotechnology Journal* 15 (2020): 1–10, https://doi.org/10.1002/biot.201900172.