Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Variable-Specific Preprocessing Enables Cross-Scale Transferable Raman Models from High-Throughput Bioreactor Data

Sung-Hyuk Han

shhan85@gccorp.com

GC Biopharma corp https://orcid.org/0009-0006-4914-8075

Jiyun Park

GC Biopharma corp

Kwang-Bae Lee

GC Biopharma corp

Cheol-Hwan Park

Sungkyunkwan University

Dong-Yup Lee

Sungkyunkwan University

Research Article

Keywords: Raman spectroscopy, Chemometric modeling, Process Analytical Technology (PAT), Cross-scale transferability, CHO cell culture, Data preprocessing

Posted Date: April 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9374625/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Bioresources and Bioprocessing on August 29th, 2026. See the published version at https://doi.org/10.1186/s40643-026-01105-5.

Variable-Specific Preprocessing Enables Cross-Scale Transferable Raman Models from High-Throughput Bioreactor Data

Sung-Hyuk Han[a,†,*], Jiyun Park[a,†], Kwang-Bae Lee[a], Cheol-Hwan Park[b], Dong-Yup Lee[b,*]

[a] Upstream Process Engineering, Manufacturing Science & Technology, R&D, GC Biopharma, 93, Ihyun-ro, 30beon-gil, Giheung-gu, Yongin-si, Gyeonggi-do 16924, Republic of Korea

[b] School of Chemical Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do 16419, Republic of Korea

[†] Sung-Hyuk Han and Jiyun Park should be considered joint first author.

[*] Corresponding authors

Email addresses: shhan85@gccorp.com (Sung-Hyuk Han), dongyuplee@skku.edu (Dong-Yup Lee).

Abstract

Raman spectroscopy is increasingly used as a process analytical technology (PAT) for real-time monitoring of bioprocesses. However, chemometric models developed using high-throughput (HT) mini-bioreactor systems often show limited predictive performance when applied to larger-scale processes, reflecting an out-of-distribution (OOD) challenge in cross-scale model transfer. In this study, we investigated whether variable-specific data preprocessing strategies can improve the cross-scale predictive performance of Raman chemometric models developed solely from HT cell culture data. Multiple preprocessing approaches were systematically evaluated for key cell culture variables, and the optimal preprocessing pipeline for each variable was selected based on its ability to minimize cross-scale prediction error. The results demonstrate that appropriate variable-specific preprocessing can substantially improve the cross-scale predictive performance of models developed from a single HT scale. Importantly, the improvement was not dependent on a specific preprocessing technique but rather on selecting preprocessing strategies suited to the spectral characteristics of each variable. These findings suggest that intelligent preprocessing selection can mitigate out-of-distribution effects in Raman chemometric models and enable more reliable prediction across scales using only HT datasets. This approach highlights the potential of Raman spectroscopy as a practical and scalable PAT tool for industrial bioprocess development.

Keywords: Raman spectroscopy; Chemometric modeling; Process Analytical Technology (PAT); Cross-scale transferability; CHO cell culture; Data preprocessing

1. Introduction

Biopharmaceutical manufacturing demands precise real-time monitoring and control to ensure product safety, efficacy, and consistent quality (Rathore & Winkle, 2009). Biologics, including recombinant proteins and monoclonal antibodies, are structurally complex and therefore require tightly regulated environments for optimal expression and production (Esmonde-White et al., 2022). Because even minor variations in critical process parameters (CPPs) can substantially affect critical quality attributes (CQAs); robust monitoring, therefore, is essential (Han et al., 2025; McAndrew & Kauffman, 2020). Process Analytical Technology (PAT) provides a framework for real-time monitoring that enhances process understanding and supports regulatory compliance (Adak et al., 2025; Food & Administration, 2004). Particularly during scale-up or site transitions, PAT tools are valuable for minimizing process variability (Esmonde-White et al., 2022). Among these tools, Raman spectroscopy has attracted growing interest for its ability to generate molecular fingerprints via inelastic light scattering. The technique operates in a non-destructive, label-free, and real-time manner, enabling simultaneous measurement of multiple variables directly in aqueous bioprocesses (Das & Agrawal, 2011; Esmonde-White et al., 2017; Li et al., 2013; Rostron et al., 2016). In sensitive cell culture systems, Raman spectroscopy enables rapid detection of environmental changes and mitigates variability during process transfer (Dong et al., 2024; Rafferty et al., 2020).

To fully leverage Raman applications, chemometric models for key variables must be developed systematically. This process involves (1) data collection, (2) spectral preprocessing, and (3) regression model construction (Gibbons et al., 2023; Otto, 2023). Because signals from target components are often weak or masked by noise and fluorescence, robust multivariate modeling is essential for accurate prediction (Dong et al., 2024; Otto, 2023; Rafferty et al., 2020; Singh, 2025). High-throughput systems, such as the Ambr® 250 HT, have been increasingly adopted with at-line Raman spectroscopy to accelerate the data collection, producing large and diverse datasets for model building concurrently with process optimization via design of experiments (Larger-Scale, 2022; Sibley et al., 2020). However, a persistent challenge arises when models developed on these mini-scale systems lose predictive accuracy

at larger scales (Rowland-Jones et al., 2021; Whelan et al., 2012). Poor cross-scale compatibility often results from cross-condition variability, in which differences in instrumentation, probe types, and spectral acquisition parameters between scales create subtantial spectral variations unrelated to chemical composition (Pétillot et al., 2020). In practical bioprocess environments, scale transitions often inherently involve differences in instrumentation, optical probes, and spectral acquisition settings; therefore, cross-scale transfer in this study reflects a composite cross-condition challenge.

Various strategies have been investigated to overcome bottlenecks, largely focusing on either expanding the calibration dataset or developing alternative modeling techniques. Several studies have focused on integrating spectra acquired at multiple scales to develop combined-scale models (Berry et al., 2015; Lang et al., 2025). While robust, this approach requires extensive large-scale datasets, diminishing the efficiency of high-throughput workflows. Other approaches seek to build models from minimal data, such as using Indirect Hard Modeling (IHM) or synthetic data augmentation (Klaverdijk et al., 2025), or by constructing models from artificially mixed samples (Hara et al., 2023). Building on these data augmentation concepts, a recent study demonstrated the generation of synthetic spectral libraries (SSL) through the in silico spiking of pure component spectra into mammalian cell culture datasets, which enriches the spectral information content while considerably reducing the empirical cost and time required for model calibration (Hellequin et al., 2025). A third category explores technically driven methods, including advanced machine learning frameworks (Rashedi et al., 2025; Tulsyan et al., 2019) or instrument standardization (Pétillot et al., 2020). However, many of these methods apply fixed methodological pipelines without systematically evaluating their suitability for specific variables. Notably, studies developing generic models have reported that certain variables, such as glutamate and titer, are inherently non-transferable under specific preprocessing schemes (Webster et al., 2018). This underscores the need to reassess spectral preprocessing and design modeling strategies tailored to each variable.

In contrast to these approaches, here, we posit that robust cross-scale and cross-condition compatibility can be achieved using data solely from a single high-throughput system. Our contribution does not lie

in proposing a new preprocessing technique, but in demonstrating that systematic, variable-specific selection of existing preprocessing strategies can substantially improve cross-scale model transferability. To realize this, we introduce a methodology that focuses on the fundamental integrity of the intrinsic spectral data itself, rather than altering the regression algorithm or the data acquisition strategy. While standard baseline correction techniques, such as Asymmetric Least Squares (ALS), are widely accessible, their universal and uniform application across all variables often fails to account for the distinct spectral characteristics, such as concentration ranges and signal overlaps, of individual metabolites. The key lies not in multi-scale datasets or reliance on a novel modeling technique, but in a systematic screening to select optimal, variable-specific preprocessing pipelines capable of isolating core chemical information from scale-dependent artifacts (e.g., background fluorescence). This addresses a critical gap in the literature: the absence of systematic, variable-specific evaluations of preprocessing as the primary enabler for single-scale model transferability, despite the recognized complexity and lack of standardization in this step (Barton et al., 2022; Gautam et al., 2015; Guo et al., 2021). To our knowledge, no previous studies have systematically demonstrated that cross-scale transferability of Raman models can be improved using only single high-throughput datasets through variable-specific preprocessing strategies. Crucially, our approach demonstrates that even widely adopted OPLS workflows can be rendered robustly transferable, provided the spectral data is properly mathematically conditioned before modeling.

In this study, we assess the transferability of Raman-based models between a high-throughput mini-bioreactor system and a bench-scale bioreactor that fundamentally involves cross-condition challenges. Specifically, the composite challenge of cross-scale transfer is addressed, which in this study inherently includes cross-instrument variability (e.g., different optical probes and detectors). We (i) benchmark the cross-scale performance limitations of conventional single-scale and combined-scale models, and (ii) demonstrate the effectiveness of our proposed methodological approach by identifying and validating variable-specific preprocessing pipelines that strengthen cross-scale robustness using only single-scale data.

2. Methods and Materials

2.1. Cell Line and Cell Culture Experiments

A stable CHO-DG44 cell line producing an enzyme was cultured in commercial chemically defined basal and feed media. A total of 30 experiments were conducted across two different scales: 7 batches in 5 L bench-scale bioreactors (3 L working volume; Eppendorf, Edison, NJ) and 23 batches in an Ambr® 250 high-throughput mini-bioreactor system (0.25 L working volume; Sartorius, Göttingen, Germany). All cultures were inoculated at an initial density of 0.2×10^6 cells/mL. To generate a diverse dataset for robust Raman modeling, a Design of Experiments (DoE) approach was implemented across the Ambr® 250 batches. This involved systematically varying key process parameters including temperature, pH, dissolved oxygen (DO), agitation, and feed strategy to capture a wide range of process variations (**Suppl. Table 1**).

2.2. Cell Culture Assay and Data Normalization

Viable cell density (VCD) and cell viability were monitored throughout the cultivation of each batch using a Cedex® HiRes Analyzer (Roche CustomBiotech, Penzberg, Germany). In addition, key metabolites including glucose, lactate, glutamine, glutamate, ammonia, and lactate dehydrogenase were analyzed using a Cedex® Bio HT analyzer (Roche CustomBiotech, Penzberg, Germany). Reference samples (Y-variables) for Raman calibration models were collected daily from the bioreactors, aligned with spectral acquisitions. Metabolite values were normalized to the [0,1] range using min-max scaling, where minimum and maximum values were determined from the dataset of each calibration model.

2.3. Raman Spectral Data Acquisition

Raman spectra were collected at regular intervals throughout the entire CHO cell culture process in darkness to prevent ambient light interference (Rowland-Jones et al., 2021). Data were acquired using two different Raman spectroscopy systems. For the Ambr® 250 HT system, an at-line BioPAT® spectro

(Sartorius, Göttingen, Germany and Endress+Hauser, Reinach, Switzerland) collected spectra via 10 scans with a 30-second exposure time per scan, while for the bench-scale bioreactor, an in-line Raman Rxn2 analyzer (Endress+Hauser, Reinach, Switzerland) acquired spectra via 75 scans with a 10-second exposure time per scan. Although the acquisition parameters differed, both systems operated with a 785 nm laser excitation wavelength, 400 mW laser power, and a spectral coverage of 100–3425 cm^{-1}. For all subsequent model development, the spectral range was limited to 450–1800 cm^{-1}.

2.4. Spectral Data Preprocessing

Prior to preprocessing, all spectra underwent a standardized quality control to remove spikes and other pronounced artifacts **(Suppl. Figure 1).** Raw Raman spectra were then preprocessed to correct for artifacts not directly related to chemical composition. All spectral preprocessing operations were systematically performed using SIMCA® software (version 17, Sartorius; Göttingen, Germany). Several established techniques were evaluated based on their common application in bioprocess modeling (Liu et al., 2021; Rafferty et al., 2020; W Eyster et al., 2021; Webster et al., 2021). Baseline variations were corrected using (1) the first derivative to remove simple additive shifts and drifts, while (2) the Asymmetric Least Squares (ALS) algorithm was employed for more complex, non-linear fluorescence backgrounds. Multiplicative effects arising from light scattering were corrected using (3) the Standard Normal Variate (SNV) transformation. The equations for these techniques are as follows:

$$y_i' = \frac{dy}{dx} \approx \frac{y_{i+1} - y_{i-1}}{2\Delta x} \quad (1)$$

where y_i' is the first derivative value at the i-th data point, y_{i+1} and y_{i-1} are values at adjacent point, and Δx is the interval between data points.

$$min \sum_{i=1}^{n} [w_i (y_i - z_i)^2 + \lambda (D^2 z)_i^2] \quad (2)$$

where y_i is the original spectral data, z_i is the estimated baseline, w_i is the weight for the residual, λ is the smoothing parameter, D^2 is the second-order difference matrix, and n is the total number of

data points.

$$y_i' = \frac{y_i - \bar{y}}{\sigma_y} \quad (3)$$

where y_i' is the normalized i-th data point, y_i is the original i-th data point, $\bar{y}$ is the mean value of the spectrum, and σ_y is its standard deviation.

2.5. Chemometric Model Development and Validation

All chemometric modeling, including the preprocessing steps described in Section 2.4 and the subsequent OPLS regression, was conducted using SIMCA® software (version 17, Sartorius; Göttingen, Germany). The entire dataset was partitioned based on its source to ensure independent external validation: the 23 Ambr® 250 batches were designated as the calibration (training) set, and the 7 bench-scale batches, which originated from a different scale, Raman instrument and culture condition, were strictly reserved as the independent validation (prediction) set. Using the training set, Orthogonal Partial Least Squares (OPLS) regression models were developed to establish a quantitative relationship between the preprocessed Raman spectra (X-variable) and the reference data (Y-variable). The optimal number of latent variables (LVs) for each model was determined, and overfitting was prevented through (4) 7-fold cross-validation. Subsequently, the final predictive accuracy of the optimized models was rigorously assessed via external validation on the independent test set. A clear distinction was maintained between internal validation metrics (R^2, Q^2, and RMSEcv) calculated from the calibration set and external validation metrics (RMSEP and bias) calculated from the independent validation set. The performance was evaluated using key metrics, including (5) the coefficient of determination (R^2), (6) the coefficient of cross-validation (Q^2), (7) root mean square error (RMSEcv and RMSEP), and (8) bias. In this study, RMSEcv and RMSEP values are expressed as relative percentages to facilitate comparison across variables with vastly different concentration ranges. The relative percentage was calculated by dividing the absolute RMSEP by the concentration range of the actual measured values for each respective variable, multiplied by 100. The equations for these indicators are as follows:

$$\text{k} - \text{fold CV error} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{i,-j})^2\right) \quad (4)$$

where n_i is the number of samples in the i-th fold, y_{ij} is the actual value of the j-th sample in the i-th fold, and $\hat{y}_{i,-j}$ is the predicted value from the model trained on all folds except the i-th fold.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (5)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value from the model, $\bar{y}$ is the mean of the actual values, and N is the total number of data points.

$$Q2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_{cv,i})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (6)$$

where $\hat{y}_{cv,i}$ is the predicted value for the i-th sample when it was held out during the cross-validation process.

$$RMSEcv\ (RSMEP) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (7)$$

where $\hat{y}_i$ is the predicted value obtained during the cross-validation process (in RMSEcv) or from the final model on the test set (for RMSEP), and N is the total number of data points.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \quad (8)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the total number of data points.

3. Results and Discussion

3.1. Single-Scale Model Performance and the Cross-Scale Compatibility Challenge

The primary objective of this study was to develop a cross-scale compatible Raman model using exclusively data from the Ambr® 250 high-throughput system. As a baseline, a 1st derivative (Der.) → SNV preprocessing pipeline was adopted as the standard approach (Otto, 2023; Singh, 2025). Following this preprocessing strategy and the OPLS modeling methodology, individual models were developed for eight key variables.

The resulting models demonstrated strong cross-validation performance within the Ambr® 250 scale (**Figure 1A**). As summarized in **Figure 2A** and **Suppl. Table 2**, all standard single-scale models exhibited low RMSEcv (%) values (3.4–8.1%), and most achieved $R^2 > 0.90$ and $Q^2 > 0.85$. The viability model showed relatively lower R^2 and Q^2 values (~0.70), likely due to the narrow dynamic range and skewed distribution of the dataset. However, despite their strong within-scale robustness indicated by low RMSEcv (%) values, several single-scale models exhibited markedly reduced predictive performance when transferred to the bench-scale bioreactor (**Figure 1A**). This deterioration was accompanied by substantially increased RMSEP (%) values and pronounced systematic biases (**Figure 2A** and **Suppl. Table 2**). Such scale-transfer discrepancies are commonly observed in chemometrics when models encounter out-of-distribution data (Guo et al., 2021). When the standard preprocessing-based single-scale (Ambr®) models were applied to real-time monitoring of the bench-scale bioreactor, predictions for VCD, glucose, and lactate showed relatively similar trends between scales. In contrast, predictions for viability, glutamine, glutamate, ammonia, and LDH exhibited substantial deviations (**Figure 2B**). This result highlights a key limitation of applying a fixed preprocessing pipeline to datasets obtained under different experimental conditions. It should be noted that these spectral deviations arise not only from volumetric scale differences but also from differences in Raman instrument optical configurations. Consequently, the results indicate that the standard preprocessing pipeline was insufficient to compensate for this compound variability.

3.2. Performance Benchmark Establishment and Key Insights into Preprocessing Strategy

To overcome the limitations of the standard single-scale model and to establish a performance benchmark, a combined-scale modeling approach was employed (A Gibbons et al., 2022; Berry et al., 2015). The model, constructed using combined Ambr® 250 and bench-scale datasets, demonstrated substantially improved predictive performance in bench-scale cultures compared with the single-scale (Ambr®-based) model (**Figure 1B**). For poorly transferred variables such as viability and glutamine, RMSEP (%) improved by factors of 1.2–3.5 (**Figure 2A**), while the bias for glutamate, ammonia, and

LDH was reduced by approximately 30% on average (**Suppl. Table 2**). Furthermore, evaluation using real-time monitoring data confirmed that the combined-scale model improved predictive performance for all variables that previously exhibited reduced prediction accuracy in the single-scale model (**Figure 2C**).

The superior performance of the combined-scale model provides insight into the underlying cause of the standard single-scale model's failure. Based on the observations described in Section 3.1, we hypothesized that the use of a universal preprocessing pipeline may be particularly disadvantageous for variables characterized by weak or highly overlapping spectral signals. In contrast to VCD, glucose, and lactate, which occur at relatively high concentrations and exhibit strong spectral features, the five problematic variables (viability, glutamine, glutamate, ammonia, and LDH) are present at lower concentrations or show limited dynamic variation. Moreover, their spectral signatures likely overlap extensively with those of background components, including complex media constituents and host cell proteins (HCPs). As a result, the analytical signals associated with these variables may be partially obscured by the surrounding spectral matrix.

Building on these observations, we formulated a mechanistic hypothesis to explain the failure of the standard single-scale model. Because the model was trained exclusively on Ambr® 250 data, it may have captured not only chemically meaningful spectral information but also scale-dependent spectral patterns, such as fluorescence baselines or impurity-related signals, that coincidentally correlated with the Y-variables within that dataset. Such correlations are likely dataset-specific rather than mechanistically meaningful and therefore may not transfer reliably to bench-scale conditions.

Within this framework, the improved performance of the combined-scale model can be interpreted as a consequence of disrupting these scale-dependent correlations. By integrating spectrally distinct datasets obtained from two different process scales (**Suppl. Figure 2**), the spectral variance of the calibration set was substantially expanded. This broader variance structure reduces the influence of spectral features that are not consistently associated with the Y-variables across scales while favoring latent structures that remain predictive in both datasets. Consequently, the model becomes more strongly

driven by chemically meaningful information shared between scales rather than by scale-dependent artifacts.

This hypothesis can be further examined by testing the inverse scenario. If spectral variance is artificially compressed, the resulting dataset should reproduce the type of information distortion observed in the single-scale model and consequently deteriorate model performance. To evaluate this effect, the order of preprocessing operations was intentionally altered by applying SNV prior to the 1st derivative. Because SNV normalizes each spectrum by its own standard deviation, both chemical signals and spectral artifacts are subjected to the same normalization process. When weak chemical signals coexist with large fluorescence baselines, the calculated standard deviation can become disproportionately influenced by these baseline contributions. Under such conditions, SNV may effectively over-normalize the spectra, compressing the overall signal space.

This signal compression is reflected in the PCA score plot (**Figure 3A**), which appears visually well aligned, and in the OPLS predictive profile (**Figure 3B**), where the dynamic range of the predictive signal is markedly distorted relative to that obtained with the standard preprocessing pipeline. Importantly, the narrowed variance observed here does not necessarily indicate a complete loss of information. Rather, it illustrates how universal preprocessing can disproportionately compress or distort variance associated with specific analytes relative to their original spectral signatures. Such structural distortion ultimately compromises the model's ability to transfer across different process scales.

Although the compressed model still shows seemingly acceptable calibration performance (low RMSEcv (%)), its reliance on distorted variance structures leads to a complete loss of predictive capability when applied to an out-of-distribution dataset, namely the bench-scale data. Even the VCD and glucose models, which exhibited low RMSEP (%) under the standard preprocessing pipeline, showed substantial performance degradation under this modified preprocessing order (**Figure 3C**). This contrast highlights that satisfactory calibration statistics alone do not guarantee robust model transferability.

Taken together, these findings suggest that the transferability of spectroscopic models is strongly influenced by how preprocessing operations shape the variance structure of the calibration dataset. Preserving predictive signal dynamics appears to be more critical for cross-scale robustness than achieving simple visual alignment of spectra.

3.3. Enhancement of Cross-Scale Compatibility via Variable-Specific Preprocessing

The insights derived from Section 3.2 highlight an inherent limitation of the combined-scale modeling approach. In the combined-scale model, spectral variance is expanded by integrating spectra obtained from different process scales, thereby exposing the model to a broader range of spectral variations and improving its robustness. Although this variance-expansion strategy proved effective, it is often impractical in real process development environments where multi-scale experimental data are limited. Therefore, the primary objective of this study was to determine whether comparable cross-scale robustness could be achieved using only single-scale data.

Rather than expanding spectral variance by incorporating spectra from additional process scales, the strategy proposed here focuses on selectively removing confounding spectral variance while preserving chemically meaningful spectral information. In particular, background fluorescence and overlapping impurity signals originating from host cell proteins (HCPs) and complex media components represent dominant sources of spectral interference in cell culture fluid (CCF) spectra. These background contributions can obscure weak Raman signals associated with several process variables and thereby hinder reliable model transferability.

To evaluate this hypothesis, multiple preprocessing pipeline combinations were systematically assessed for all eight variables (**Figure 4**). The cross-scale predictive performance of each optimized single-scale model, expressed as RMSEP (%), was compared with that of the standard single-scale model and the combined-scale benchmark model (**Figure 5)**.

The results showed that the optimal preprocessing strategy depended strongly on the spectral

characteristics of each variable. For variables characterized by strong and well-resolved Raman signals, such as VCD, glucose, and lactate, the standard preprocessing pipeline (1st Der. → SNV) already provided robust predictive performance. Additional preprocessing steps did not improve prediction accuracy and in some cases even degraded model performance (**Table 1**). This observation suggests that, for variables with strong spectral signatures, the major spectral interferences, such as baseline shifts and scattering effects, can already be adequately corrected by the standard preprocessing pipeline. In particular, for variables exhibiting broad spectral features, such as VCD, additional baseline correction may result in overcorrection if parts of the true analytical signal are incorrectly treated as baseline.

In contrast, five variables (viability, glutamine, glutamate, ammonia, and LDH) exhibited relatively poor external predictive performance despite acceptable internal validation results when the standard preprocessing pipeline was applied. These variables are characterized by weak or highly overlapping Raman signals, where complex fluorescence baselines represent a dominant source of spectral interference. When asymmetric least squares (ALS) baseline correction was incorporated into the preprocessing pipeline, RMSEP (%) for these variables improved significantly (**Figure 5A** and **5B**).

The improvement observed after introducing ALS can be attributed to its ability to effectively model and remove complex baseline variations arising from fluorescence backgrounds. By reducing these large-scale baseline contributions, ALS improves the detectability of weak chemical signals that would otherwise remain partially masked within the fluorescence background.

The impact of ALS extends beyond simple baseline correction. Model bias frequently arises when spurious correlations are formed between systematic background variations, such as fluorescence intensity, and analyte concentrations within the calibration dataset. As illustrated in **Figure 5C**, the introduction of ALS substantially reduced model bias. By suppressing fluctuating background signals, ALS limits the formation of such non-causal correlations and stabilizes the covariance structure between the spectral matrix (X) and the reference variables (Y).

From a chemometric perspective, this effect can be interpreted as a restructuring of the X-variance space

prior to PLS modeling. By reducing variance components associated with fluorescence background and other non-informative spectral contributions, ALS effectively increases the relative contribution of chemically relevant spectral variance. Consequently, the latent variables extracted by the PLS algorithm are more likely to capture analyte-related spectral information rather than scale-dependent spectral artifacts (Gautam et al., 2015; Rinnan et al., 2009).

It should be emphasized that ALS itself is not a newly proposed preprocessing method but a well-established technique widely used in chemometrics-based Raman analysis. However, the results of this study demonstrate that the effectiveness of such preprocessing approaches strongly depends on the spectral characteristics of the target variable. For variables such as VCD, glucose, and lactate, which occur at relatively high concentrations in cell culture fluid and exhibit strong and well-resolved Raman features, the application of ALS provided little additional benefit (**Table 1**). In contrast, variables characterized by weak or overlapping spectral signatures showed clear improvements when ALS was incorporated into the preprocessing pipeline. These observations highlight that preprocessing strategies should not be applied uniformly across all analytes but rather tailored according to the spectral characteristics of individual variables. In this context, ALS represents one example of a variable-specific preprocessing strategy capable of improving cross-scale model robustness.

These findings were further supported by OPLS-based diagnostic analysis (**Table 2**). External validation data obtained at the bench scale were projected onto both the standard single-scale model and the optimized models to examine the relationship between predictive performance and spectral variance diagnostics. The optimized models consistently showed substantial improvements in predictive accuracy (RMSEP, **Figure 5B**) and bias (**Figure 5C**) compared with the standard model. However, the S/N ratio derived from X-variance diagnostics exhibited variable-dependent behavior. As shown in **Table 2**, the S/N ratio decreased for viability and glutamate, increased for glutamine, and remained largely unchanged for ammonia and LDH. These results suggest that X-variance-based diagnostic indicators alone may not provide a consistent interpretation when evaluating cross-scale model transferability. Because models handle scale-dependent spectral artifacts differently depending

on the variable, this complexity cannot be fully captured by simple X-variance metrics. Therefore, external Y-based validation metrics, particularly RMSEP and bias, provide a more direct and reliable basis for evaluating cross-scale model robustness.

To further verify the statistical validity of the models, permutation tests (n = 100) were performed. The original Q^2 values (**Figure 2A**) calculated from the true datasets were significantly positive, whereas models constructed using randomly permuted Y-data consistently produced negative Q^2 values (**Figure 5D**). These results confirm that the predictive performance of both the standard and optimized models was statistically valid and not the result of chance fitting. Although the external validation dataset consisted of seven independent batches, the statistical significance observed in the paired t-tests (**Figure 5B**) and permutation tests indicate that the observed improvements cannot be explained solely by sample size effects.

Overall, single-scale models constructed using optimized variable-specific preprocessing pipelines achieved substantially improved predictive accuracy compared with models built using the standard preprocessing strategy and, in several cases, approached the performance of the resource-intensive combined-scale benchmark model. These findings demonstrate that selecting preprocessing strategies according to the spectral characteristics of individual variables is a critical determinant of cross-scale Raman model transferability. From a practical perspective, the present results indicate that robust cross-scale Raman models can be developed using only high-throughput mini-bioreactor data when appropriate variable-specific preprocessing strategies are applied. Such an approach may substantially reduce the reliance on large-scale experiments during PAT model development and provide a cost-efficient and agile modeling strategy for modern biopharmaceutical process development.

4. Conclusion

This study addresses the critical challenge of cross-scale compatibility in Raman-based monitoring of CHO cell culture processes. Standard single-scale models showed limited transferability between

process scales, whereas combined-scale models improved prediction performance but required additional multi-scale experimental data, limiting their practicality during early-stage development.

By strategically optimizing the spectral preprocessing pipeline, we demonstrate that models trained exclusively on mini-scale data can achieve cross-scale predictive performance approaching that of combined-scale benchmark models. Importantly, the optimal preprocessing strategy was variable-dependent and determined by the interaction between Raman signal characteristics and dominant spectral interferences in the cell culture matrix. Rather than applying a universal preprocessing pipeline, selecting preprocessing methods that preserve analyte-related spectral variance proved critical for improving model transferability.

From a practical perspective, this approach enables the development of robust and transferable Raman-based PAT models using only high-throughput mini-bioreactor systems, thereby reducing the reliance on resource-intensive large-scale experiments during process development. Within the broader analytical framework for Raman-based bioprocess monitoring (Guo et al., 2021), this strategy provides a rapid and experimentally efficient alternative to approaches based on culture-free chemical mixtures (Hara et al., 2023). By enabling earlier and more reliable implementation of spectroscopic monitoring, the methodology supports Quality-by-Design (QbD) principles and facilitates more agile biomanufacturing processes.

Future work will focus on integrating the optimized variable-specific preprocessing strategies identified here with formal model transfer techniques (Barton et al., 2022) to further enhance long-term robustness and scalability of Raman-based monitoring models across bioprocess platforms.

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to [security regulations of GC Biopharma] but are available from the corresponding author on reasonable request.

Completing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Technology Innovation Program funded by the Ministry of Trade, Industry and Resources (MOTIR, Korea) (Development of a digital twin linked continuous culture-based bioreactor production system for the continuous manufacturing of recombinant protein therapeutics for rare disease treatment, Project No. 2410010642, RS-2024-004316) and by GC Biopharma.

Authors' contributions

Sung-Hyuk Han: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review and editing, Supervision, Project administration, Funding acquisition **Jiyun Park:** Conceptualization, Methodology, Formal analysis, Investigation, Software, Data curation, Writing – review and editing, Visualization. **Kwangbae Lee:** Methodology, Formal analysis, Data

curation, Resources. **Cheol-Hwan Park:** Investigation, Resources. **Dong-Yup Lee:** Supervision.

Acknowledgements

Not applicable

References

A Gibbons, L., Rafferty, C., Robinson, K., Abad, M., Maslanka, F., Le, N., Mo, J., Clark, K., Madden, F., & Hayes, R. (2022). Raman based chemometric model development for glycation and glycosylation real time monitoring in a manufacturing scale CHO cell bioreactor process. *Biotechnology Progress*, *38*(2), e3223. https://doi.org/10.1002/btpr.3223

Adak, R., Jana, A., Debnath, B., Das, A., & Bandyopadhyay, R. (2025). A Review on Process Analytical Technology as a Driver of Pharmaceutical Manufacturing for the Improvement of Quality while Reducing Costs. *Current Drug Safety*. https://doi.org/10.2174/0115748863374678250501184523

Barton, B., Thomson, J., Diz, E. L., & Portela, R. (2022). Chemometrics for Raman spectroscopy harmonization. *Applied spectroscopy*, *76*(9), 1021-1041. https://doi.org/10.1177/00037028221094070

Berry, B., Moretto, J., Matthews, T., Smelko, J., & Wiltberger, K. (2015). Cross-scale predictive modeling of CHO cell culture growth and metabolites using R aman spectroscopy and multivariate analysis. *Biotechnology Progress*, *31*(2), 566-577. https://doi.org/10.1002/btpr.2035

Das, R. S., & Agrawal, Y. K. (2011). Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*, *57*(2), 163-176. https://doi.org/10.1016/j.vibspec.2011.08.003

Dong, X., Yan, X., & Qu, H. (2024). Improving Raman-Based Models for Real-Time Monitoring the

CHO Cell Culture Process with Effective Variable Selection Strategies. *Applied Sciences*, *14*(19), 8890. https://doi.org/10.3390/app14198890

Esmonde-White, K. A., Cuellar, M., & Lewis, I. R. (2022). The role of Raman spectroscopy in biopharmaceuticals from development to manufacturing. *Analytical and bioanalytical chemistry*, 1-23. https://doi.org/10.1007/s00216-021-03727-4

Esmonde-White, K. A., Cuellar, M., Uerpmann, C., Lenain, B., & Lewis, I. R. (2017). Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Analytical and bioanalytical chemistry*, *409*(3), 637-649. https://doi.org/10.1007/s00216-016-9824-1

Food, & Administration, D. (2004). Guidance for industry, PAT-A framework for innovative pharmaceutical development, manufacturing and quality assurance. *http://www. fda. gov/cder/guidance/published. html*.

Gautam, R., Vanga, S., Ariese, F., & Umapathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, *2*(1), 8. https://doi.org/10.1140/epjti/s40485-015-0018-6

Gibbons, L., Maslanka, F., Le, N., Magill, A., Singh, P., Mclaughlin, J., Madden, F., Hayes, R., McCarthy, B., & Rode, C. (2023). An assessment of the impact of Raman based glucose feedback control on CHO cell bioreactor process development. *Biotechnology Progress*, *39*(5), e3371. https://doi.org/10.1002/btpr.3371

Guo, S., Popp, J., & Bocklitz, T. (2021). Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling. *Nature protocols*, *16*(12), 5426-5459. https://doi.org/10.1038/s41596-021-00620-3

Han, S.-H., Park, S.-Y., Cha, H.-M., Lee, K.-B., Lim, J.-H., & Lee, D.-Y. (2025). A robust scale-down model development and process characterization for monoclonal antibody biomanufacturing using multivariate data analysis. *Journal of biotechnology*, *401*, 11-20. https://doi.org/10.1016/j.jbiotec.2025.02.007

Hara, R., Kobayashi, W., Yamanaka, H., Murayama, K., Shimoda, S., & Ozaki, Y. (2023). Development

of Raman calibration model without culture data for in-line analysis of metabolites in cell culture media. *Applied spectroscopy*, *77*(5), 521-533. https://doi.org/10.1177/00037028231160197

Hellequin, L. V., Borràs, V. J., Romann, P., Vishwanathan, N., Souquet, J., & Villiger, T. K. (2025). Synthetic spectral libraries for Raman model calibration. *Analytical and bioanalytical chemistry*, *417*(25), 5675-5689. https://doi.org/10.1007/s00216-025-05985-y

Klaverdijk, M., Smulders, L. A., Ottens, M., & Klijn, M. E. (2025). Towards Rapid Calibration of Bioprocess Quantification Models Using Single Compound Raman Spectra: A Comparison of Four Approaches. *Biotechnology and bioengineering*. https://doi.org/10.1002/bit.70092

Lang, Z., Chen, G., Yan, S., Zhang, Z., Yang, Y., Tang, Z., Zhu, H., Dong, S., Zhou, H., & Zhou, W. (2025). Enhancing cross-scale Raman in-line monitoring capability of cell culture process in large-scale manufacturing. *AIChE Journal*, *71*(1), e18608. https://doi.org/10.1002/aic.18608

Larger-Scale, S.-T. B. (2022). Raman Spectrometric PAT Models. *BioProcess International*, *20*, 9.

Li, B., Ray, B. H., Leister, K. J., & Ryder, A. G. (2013). Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. *Analytica chimica acta*, *796*, 84-91. https://doi.org/10.1016/j.aca.2013.07.058

Liu, Z., Zhang, Z., Qin, Y., Chen, G., Hu, J., Wang, Q., & Zhou, W. (2021). The application of Raman spectroscopy for monitoring product quality attributes in perfusion cell culture. *Biochemical Engineering Journal*, *173*, 108064. https://doi.org/10.1016/j.bej.2021.108064

McAndrew, J., & Kauffman, G. (2020). Upstream Process Development with ak La Criterion Within the BIOne Single-Use Bioreactor [White paper]. Distek, Inc. In.

Otto, M. (2023). *Chemometrics: statistics and computer application in analytical chemistry*. John Wiley & Sons.

Pétillot, L., Pewny, F., Wolf, M., Sanchez, C., Thomas, F., Sarrazin, J., Fauland, K., Katinger, H., Javalet, C., & Bonneville, C. (2020). Calibration transfer for bioprocess Raman monitoring using Kennard Stone piecewise direct standardization and multivariate algorithms. *Engineering Reports*, *2*(11), e12230. https://doi.org/10.1002/eng2.12230

Rafferty, C., Johnson, K., O'Mahony, J., Burgoyne, B., Rea, R., & Balss, K. M. (2020). Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. *Biotechnology Progress*, *36*(4), e2977. https://doi.org/10.1002/btpr.2977

Rashedi, M., Demers, M., Khodabandehlou, H., Wang, T., Garvin, C., & Rianna, S. (2025). Continuous glucose feedback control using Raman spectroscopy and deep learning models for biopharmaceutical processes. *Biotechnology Progress*, *41*(4), e70020. https://doi.org/10.1002/btpr.70020

Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. *Nature biotechnology*, *27*(1), 26-34. https://doi.org/10.1038/nbt0109-26

Rinnan, Å., Berg, F. v. d., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201-1222. https://doi.org/10.1016/j.trac.2009.07.007

Rostron, P., Gaber, S., & Gaber, D. (2016). Raman spectroscopy, review. *laser*, *21*, 24.

Rowland-Jones, R. C., Graf, A., Woodhams, A., Diaz-Fernandez, P., Warr, S., Soeldner, R., Finka, G., & Hoehse, M. (2021). Spectroscopy integration to miniature bioreactors and large scale production bioreactors–Increasing current capabilities and model transfer. *Biotechnology Progress*, *37*(1), e3074. https://doi.org/10.1002/btpr.3074

Sibley, M., Woodhams, A., Hoehse, M., & Zoro, B. (2020). Novel integrated Raman spectroscopy Technology for Minibioreactors. *BioProcess Int*, *18*, 9.

Singh, V. K. (2025). *Applied Raman Spectroscopy: Concepts, Instrumentation, Chemometrics, and Life Science Applications*. Elsevier.

Tulsyan, A., Schorner, G., Khodabandehlou, H., Wang, T., Coufal, M., & Undey, C. (2019). A machine-learning approach to calibrate generic Raman models for real-time monitoring of cell culture processes. *Biotechnology and bioengineering*, *116*(10), 2575-2586. https://doi.org/10.1002/bit.27100

W Eyster, T., Talwar, S., Fernandez, J., Foster, S., Hayes, J., Allen, R., Reidinger, S., Wan, B., Ji, X., & Aon, J. (2021). Tuning monoclonal antibody galactosylation using Raman spectroscopy-

controlled lactic acid feeding. *Biotechnology Progress*, *37*(1), e3085. https://doi.org/10.1002/btpr.3085

Webster, T. A., Hadley, B. C., Dickson, M., Busa, J. K., Jaques, C., & Mason, C. (2021). Feedback control of two supplemental feeds during fed-batch culture on a platform process using inline Raman models for glucose and phenylalanine concentration. *Bioprocess and Biosystems Engineering*, *44*(1), 127-140. https://doi.org/10.1007/s00449-020-02429-y

Webster, T. A., Hadley, B. C., Hilliard, W., Jaques, C., & Mason, C. (2018). Development of generic raman models for a GS-KOTM CHO platform process. *Biotechnology Progress*, *34*(3), 730-737. https://doi.org/10.1002/btpr.2633

Whelan, J., Craven, S., & Glennon, B. (2012). In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnology Progress*, *28*(5), 1355-1362. https://doi.org/10.1002/btpr.1590

Tables

Table 1. Preprocessing pipeline screening candidates and selection for optimal variable-specific Raman models. This table outlines the various data preprocessing pipelines evaluated for eight key variables (VCD, viability, glucose, lactate, glutamine, glutamate, ammonia, and LDH). The pipelines combined techniques including the first derivative (1st D), standard normal variate (S), and asymmetric least squares (A). The optimal pipeline for each variable was selected based on its ability to minimize the root mean square error of prediction (RMSEP, %) across scales. For variables with strong spectral signatures, the standard pipeline (1st D → S) was optimal, whereas complex variables required additional baseline correction.

	RMSEP, %				Optimal Pipeline Selected	ΔRMSEP (%)
	Std. pipeline	Screening Pipeline Candidates				
	1st D → S	1st D → S → A	A → 1st D → S	S → 1st D → A		
VCD	6.6	12.9	7.4	22.2	1st D → S	N/A
Viability	15.8	10.0	11.6	17.5	1st D → S → A	36.7
Glucose	5.3	7.4	6.0	29.3	1st D → S	N/A
Lactate	6.4	9.5	6.8	5.8	1st D → S	N/A
Glutamine	18.0	11.7	13.3	16.5	1st D → S → A	35.0
Glutamate	13.5	10.6	7.3	50.7	A → 1st D → S	45.9
Ammonia	10.7	9.2	15.3	53.8	1st D → S → A	14.0
LDH	10.7	6.0	4.7	19.4	A → 1st D → S	56.1

Table 2. Orthogonal Partial Least Squares (OPLS) diagnostic analysis of spectral variance and signal-to-noise ratio. OPLS-based diagnostic analysis comparing the standard preprocessing pipeline with the optimized variable-specific pipelines. The table presents the variance explained by the predictive component, the orthogonal component, and the resulting signal-to-noise (S/N) ratio. The variable-dependent behavior of the S/N ratio highlights that external validation metrics, such as RMSEP and bias, are more reliable indicators of cross-scale robustness than X-variance-based diagnostics alone.

	Variance Explained by Predictive Comp. (R2X[1])		Variance Explained by Orthogonal Comp. (R2Xo[1])		Signal to Noise (S/N)	
	Std. Pipeline	Opt. Pipeline	Std. Pipeline	Opt. Pipeline	Std. Pipeline	Opt. Pipeline
Viability	0.341	0.307	0.225	0.357	1.34	0.86
Glutamine	0.242	0.238	0.279	0.234	0.87	1.02
Glutamate	0.498	0.478	0.113	0.117	4.41	4.09
Ammonia	0.434	0.450	0.142	0.149	3.06	3.02
LDH	0.293	0.290	0.191	0.187	1.53	1.55

Figure caption

Figure 1. Performance comparison between standard single-scale and combined-scale models for cross-scale monitoring. Cross-validation and external prediction performance of Raman models. (A) The single-scale model demonstrates internal cross-validation (blue circles) but exhibits poor external prediction when applied to bench-top cultures (red triangles). (B) The combined-scale model displays cross-validation (green circles) and significantly improved external prediction performance for bench-top cultures (red triangles) compared to the single-scale approach.

Figure 2. Predictive performance metrics and real-time monitoring evaluation of single-scale and combined-scale models. Quantitative assessment of model transferability to bench-top systems. (A) Evaluation metrics comparing the single-scale model (pink circles) and the combined-scale model (navy circles). Gray arrows highlight the substantial improvement in RMSEP (%) achieved by the combined-scale model. (B) Real-time bench-top monitoring using the single-scale model (gray line) versus offline reference measurements (red circles and triangles), revealing large predictive gaps (n=2 batches). (C) Real-time bench-top monitoring using the combined-scale model (gray line) alongside offline reference measurements (red circles and triangles). The model demonstrates high predictive accuracy aligned with actual values (n=2 batches).

Figure 3. Effect of altered preprocessing order on spectral variance and cross-scale model transferability. Impact of standard normal variate (SNV) application timing. (A) PCA score plots of the calibration dataset comparing the standard 1st Der. → SNV pipeline (blue circles) against the SNV → 1st Der. pipeline (red triangles). The reverse pipeline artificially clusters the data structure around the origin. (B) OPLS Y-related predictive profiles for independent bench-scale data. The model using the standard pipeline (left Y-axis, blue line) maintains a full dynamic range. Conversely, the reverse pipeline (right Y-axis, red line) exhibits a severely compressed dynamic range, indicating a distortion of predictive information. Apparent visual data harmonization in PCA does not ensure predictive robustness. (C) External validation performance comparison showing that altering the preprocessing order does not impact internal validation but drastically increases external prediction errors (standard

preprocessing: blue bar; reverse preprocessing: red bar).

Figure 4. Schematic workflow of variable-specific preprocessing pipelines for single-scale Raman model optimization. Systematic screening flow for variable-specific preprocessing. First, a specific spectral region (400 to 1800 cm^{-1}) is selected from the full raw Raman spectra. Next, four distinct preprocessing pipelines are applied to the spectral data. Finally, the RMSEP is evaluated for eight key variables to select the single optimal model demonstrating the lowest prediction error (red line) for each specific parameter.

Figure 5. Cross-scale predictive performance of single-scale models using optimized variable-specific preprocessing pipelines. Comparison of predictive capabilities among standard single-scale, combined-scale, and optimized single-scale models. (A) External prediction error (RMSEP, %) of the standard single-scale (sky blue), combined-scale (gray), and optimized single-scale (blue) models. (B) Paired t-test results for internal (RMSEcv) and external (RMSEP) errors across eight variables. All three models show no statistical difference in internal errors, whereas the standard single-scale model exhibits statistically higher external errors than the other two. (C) Prediction bias for five variables with weak signal characteristics. The optimized single-scale model effectively reduces bias across all parameters compared to the standard model. (D) Permutation tests (100 permutations) for the five weak-signal variables (viability, glutamine, glutamate, ammonia, LDH) using the standard (blue dots) and optimized (green dots) single-scale models. All models yield a negative Q2 from randomly permuted Y-data, confirming statistical validity without chance correlation.

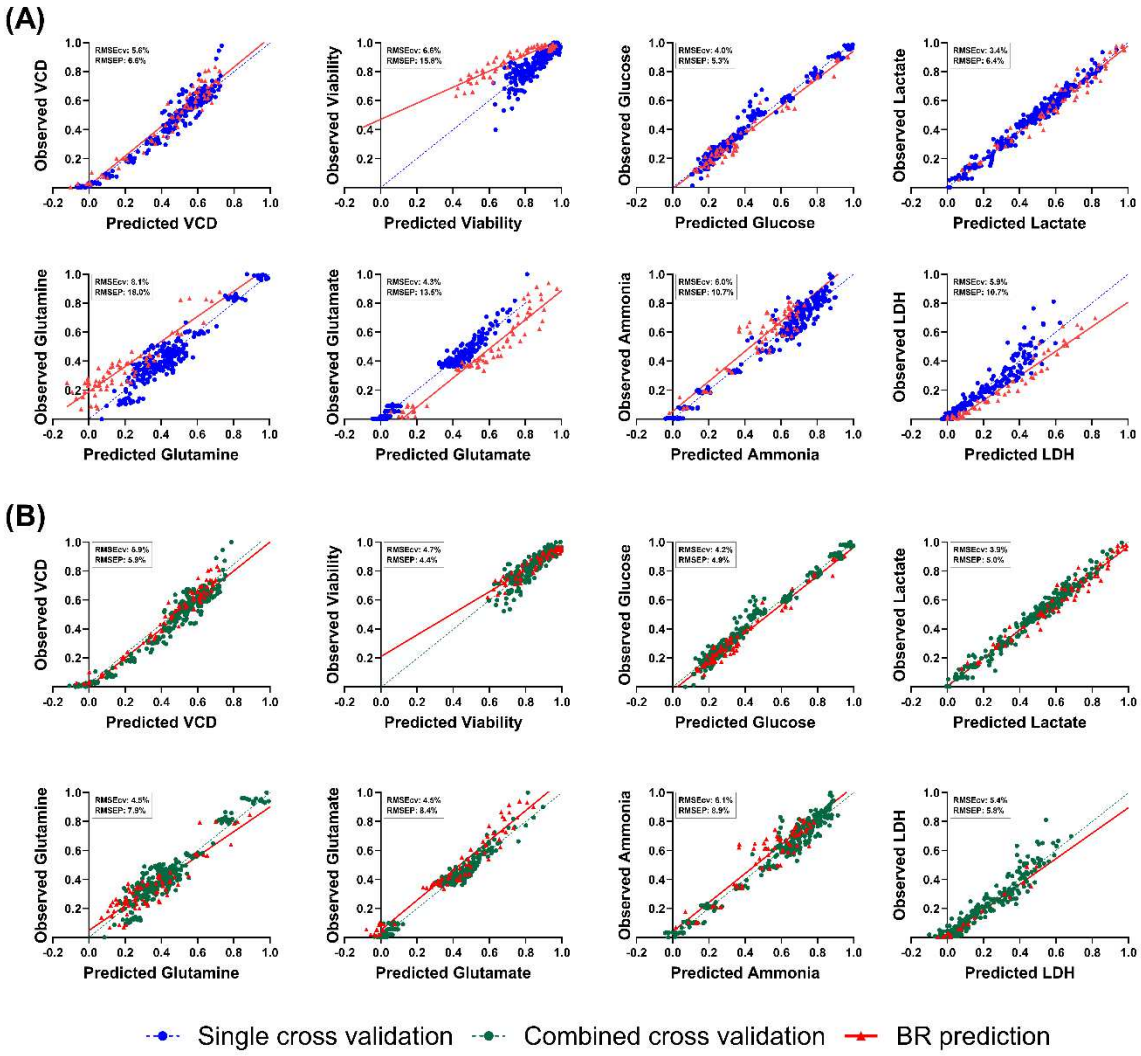

(A)
Observed VCD
Predicted VCD
RMSEcv: 5.6%
RMSEP: 6.6%
Observed Viability
Predicted Viability
RMSEcv: 6.6%
RMSEP: 15.8%
Observed Glucose
Predicted Glucose
RMSEcv: 4.0%
RMSEP: 5.3%
Observed Lactate
Predicted Lactate
RMSEcv: 3.4%
RMSEP: 6.4%
Observed Glutamine
Predicted Glutamine
RMSEcv: 8.1%
RMSEP: 18.0%
Observed Glutamate
Predicted Glutamate
RMSEcv: 4.3%
RMSEP: 13.6%
Observed Ammonia
Predicted Ammonia
RMSEcv: 6.0%
RMSEP: 10.7%
Observed LDH
Predicted LDH
RMSEcv: 5.9%
RMSEP: 10.7%
(B)
Observed VCD
Predicted VCD
RMSEcv: 5.9%
RMSEP: 5.8%
Observed Viability
Predicted Viability
RMSEcv: 4.7%
RMSEP: 4.4%
Observed Glucose
Predicted Glucose
RMSEcv: 4.2%
RMSEP: 4.9%
Observed Lactate
Predicted Lactate
RMSEcv: 3.9%
RMSEP: 5.0%
Observed Glutamine
Predicted Glutamine
RMSEcv: 4.5%
RMSEP: 7.9%
Observed Glutamate
Predicted Glutamate
RMSEcv: 4.6%
RMSEP: 8.4%
Observed Ammonia
Predicted Ammonia
RMSEcv: 6.1%
RMSEP: 6.9%
Observed LDH
Predicted LDH
RMSEcv: 5.4%
RMSEP: 5.8%
Single cross validation
Combined cross validation
BR prediction

Figure 1

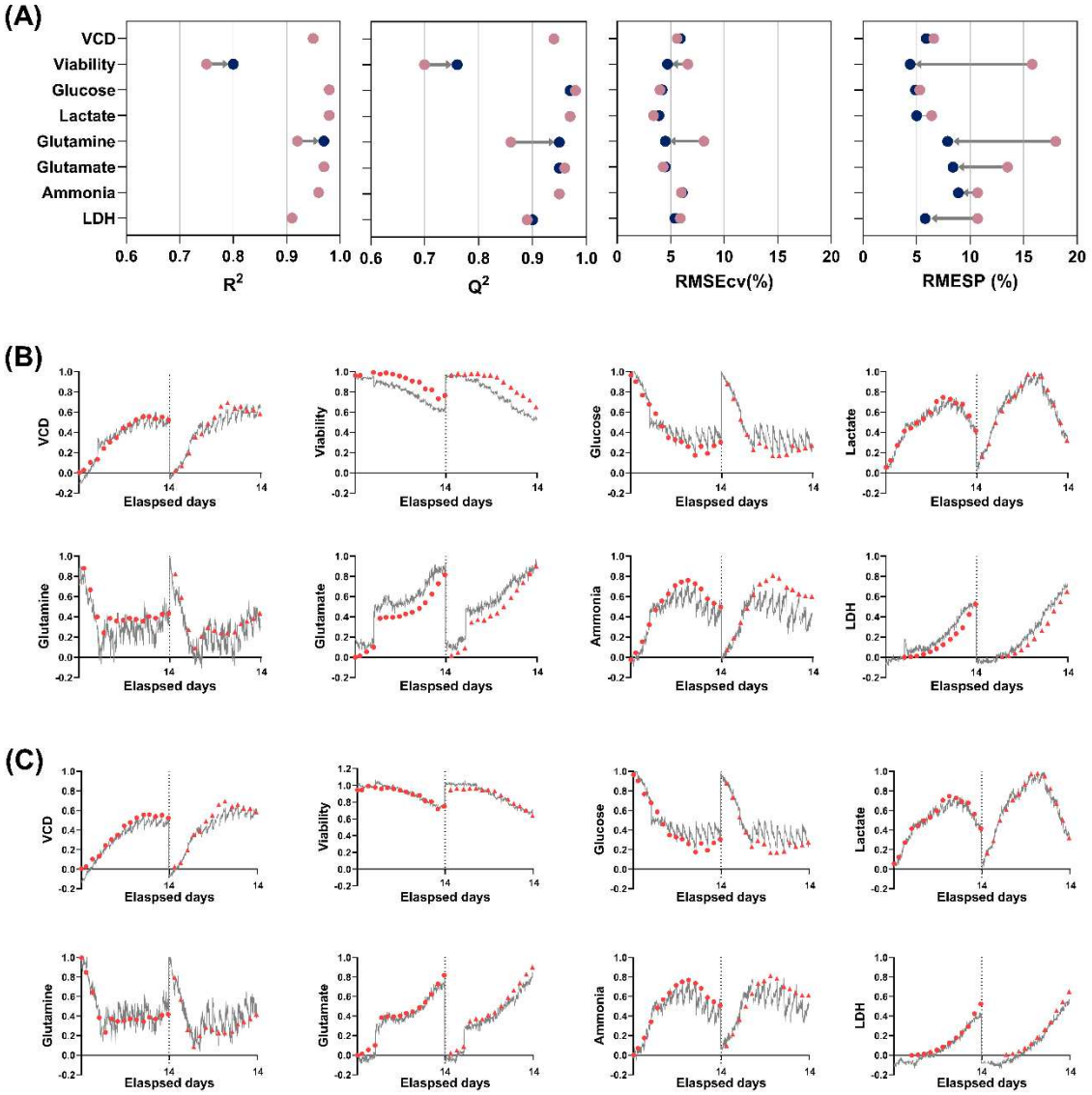
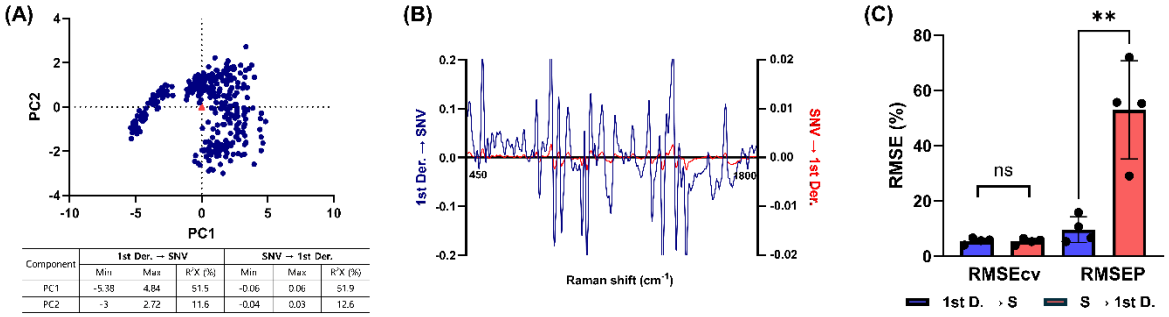

(A)
VCD
Viability
Glucose
Lactate
Glutamine
Glutamate
Ammonia
LDH
R²
Q²
RMSEcv(%)
RMESP (%)
(B)
(C)
Elaspsed days

Figure 2



Component	1st Der. → SNV			SNV → 1st Der.		
	Min	Max	R²X (%)	Min	Max	R²X (%)
PC1	-5.38	4.84	51.5	-0.06	0.06	51.9
PC2	-3	2.72	11.6	-0.04	0.03	12.6

Figure 3

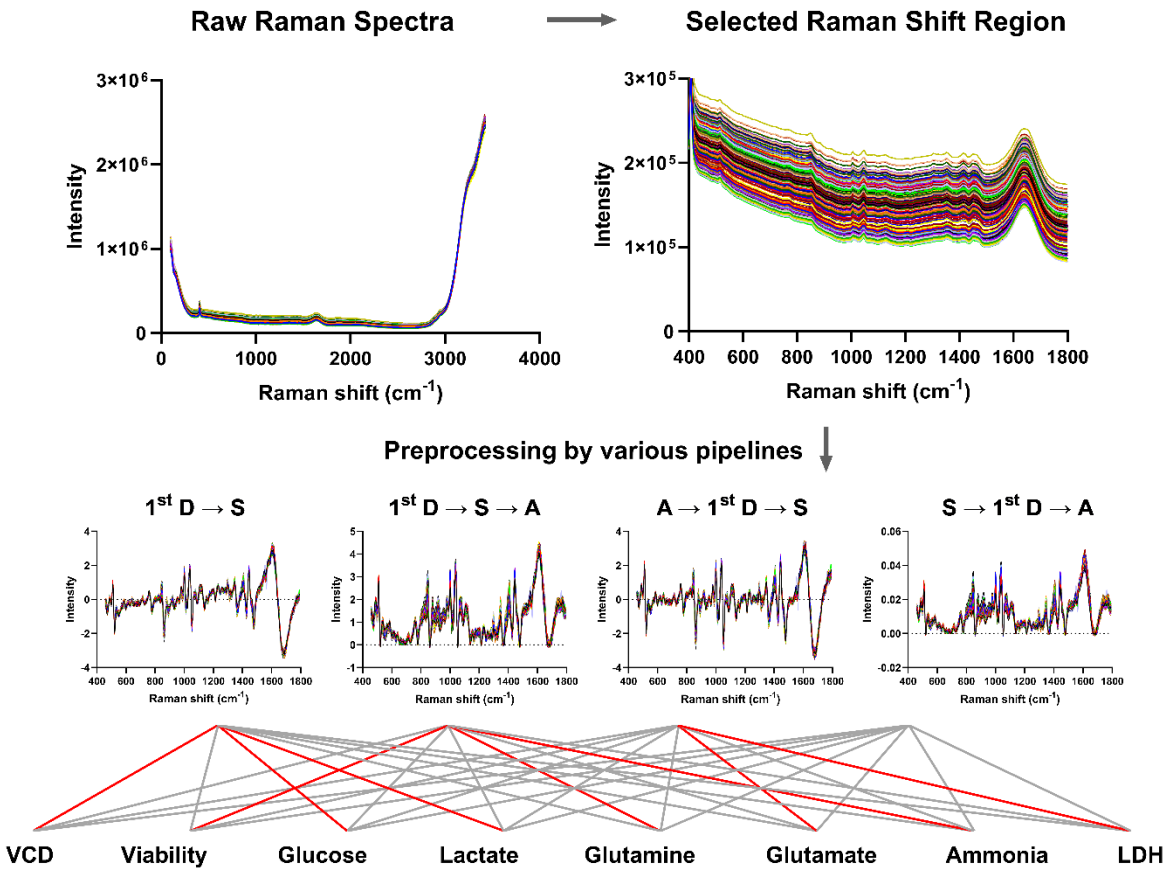
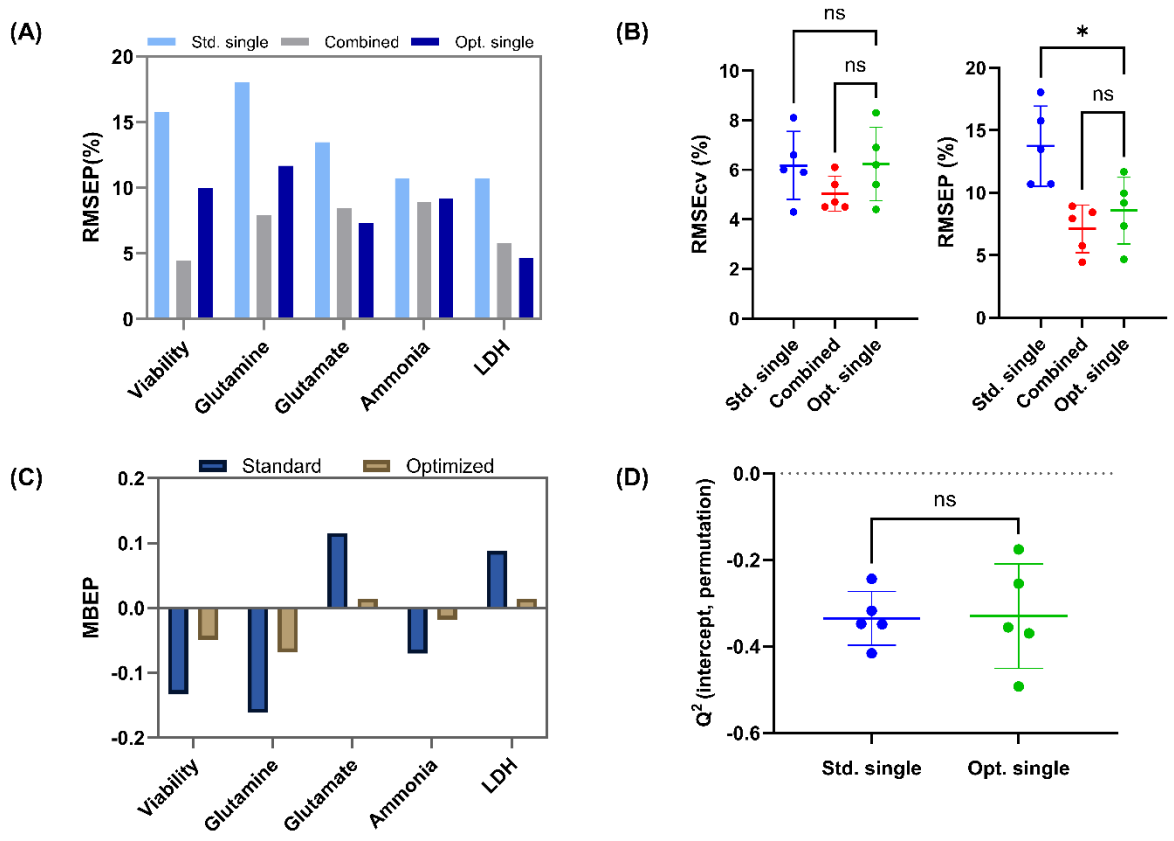

Raw Raman Spectra
Selected Raman Shift Region
Intensity
Raman shift (cm-1)
Preprocessing by various pipelines
1st D → S
1st D → S → A
A → 1st D → S
S → 1st D → A
VCD
Viability
Glucose
Lactate
Glutamine
Glutamate
Ammonia
LDH

Figure 4



704

705 Figure 5

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstractBIOB.jpg
- SupplementaryTablesandFiguresBIOB.docx