
Gianmarco Luchetti Sfondalmondo

Marieke E. Klijn

Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, Delft, 2629 HZ, The Netherlands

g.luchettisfondalmondo@tudelft.nl

Balancing prediction performance and spectral interpretability: Multi-metric guided design of Raman spectrum preprocessing pipelines to transfer single compound models to bioprocess data

Raman spectroscopy is a widely explored in-line process analytical technology due to its high chemical specificity and suitability for chemometric modelling [1]. However, in the field of biotechnology, developing robust chemometric models for Raman spectroscopy-based bioprocess monitoring typically requires extensive bioreactor runs for data collection. Single-compound calibration is a cost-effective alternative, but spectral preprocessing is necessary for the transfer of models calibrated on abiotic spectra (i.e., an analyte in a fermentation medium) to complex bioprocess spectra [2]. In this work, we present a multi-metric framework for systematic preprocessing selection. This was used to screen ~65,000 preprocessing combinations, where each pipeline was evaluated across signal preservation, noise reduction, peak resolution, and model performance. Selected pipelines were applied to preprocess single-compound spectra for Partial Least Squares (PLS) calibration. The same pipelines were subsequently used to preprocess bioprocess spectra for prediction of ethanol and glucose concentration, evaluating directly model transfer to four batch and one fed-batch aerobic *Saccharomyces cerevisiae* fermentation datasets. Compared to preprocessing pipelines selected based solely on prediction error, the multi-metric framework yielded more robust and chemically interpretable models. Using a single latent variable PLS, the ethanol model performed within ± 2 percentage points of error respect to literature benchmarks, demonstrating that preprocessing optimization for compounds with intense, narrow peaks can successfully transfer from abiotic to complex bioprocess matrices. Moreover, this was achieved without a process-specific internal standard or biomass spectra, indicating flexibility and applicability to other processes. However, inherent weaker spectral features of glucose reduced model performance, highlighting the limitations of a single-compound model for these types of compounds. Our findings demonstrate the value of analyte-specific preprocessing strategies and how quantitative multi-objective preprocessing selection can support efficient development of cost-effective, transferable models.

References

- [1] K. A. Esmonde-White, M. Cuellar, and I. R. Lewis, "The role of Raman spectroscopy in biopharmaceuticals from development to manufacturing," *Anal. Bioanal. Chem.*, vol. 414, no. 2, pp. 969–991, Jan. 2022, doi: 10.1007/s00216-021-03727-4.
- [2] M. Klaverdijk, L. A. Smulders, M. Ottens, and M. E. Klijn, "Towards Rapid Calibration of Bioprocess Quantification Models Using Single Compound Raman Spectra: A Comparison of Four Approaches," *Biotechnol. Bioeng.*, vol. 123, no. 2, pp. 324–336, Feb. 2026, doi: 10.1002/bit.70092.