

# Chemometric Models Applied to Raman Spectroscopy for Bioprocess Monitoring

Silvia Lisci<sup>a</sup>, Laura Mais<sup>a</sup>, Alice Corda<sup>a</sup>, Stefania Tronci<sup>a\*</sup>, Massimiliano Errico<sup>b</sup>, Massimiliano Grosso<sup>a</sup>

<sup>a</sup>Dipartimento di Ingegneria Meccanica, Chimica e dei Materiali, Università degli Studi di Cagliari, Cagliari I-09123, Italy

<sup>b</sup>University of Southern Denmark, Faculty of Engineering, Department of Green Technology, DK-5230 Odense M, Denmark  
[stefania.tronci@unica.it](mailto:stefania.tronci@unica.it)

This work investigates the development of a Raman spectroscopy-based sensor for the quantitative analysis of products obtained from the fermentation to ethanol of the hydrolyzed product of brewer's spent grain. Partial least squares (PLS) and variable importance in projection (VIP) methods were applied in order to correlate the information contained in the Raman spectra to the concentration of the compounds investigated. Results showed good prediction capabilities of the developed models when analyzing standard solutions containing ethanol and glucose. On the other hand, when analyzing samples obtained from the fermentation process, turbidity, caused by suspended particles in the solution, made analyzing spectra more difficult, leading to a decrease of the validation determination coefficient from 0.9374 to 0.7913 for ethanol and from 0.6237 to 0.4783 for glucose.

## 1. Introduction

As of today, routine monitoring and control of fermentation bioreactors is limited to basic parameters, such as temperature, pressure, pH, and DO, for which there are established in situ sensor technologies (Mei et al., 2017). However, most chemical and biochemical properties are measured off-line, resulting in measurements unsuitable for process control due to the delay between sampling and analysis time (Lisci et al, 2020). In order to overcome this challenge, extensive research has been conducted over the past decade to develop sensors for monitoring bioprocesses in real time, primarily based on optical methods. Among the emerging techniques, Raman spectroscopy represents a promising methodology for monitoring biotechnological systems due to important features such as speed, high signal-to-noise ratio, good resolution, ability to provide a stable signal, and low water interference (Ávila et al., 2012; Claßen et al., 2017). Moreover, it is non-destructive and does not require sample pre-treatment (Esmonde-White et al., 2017; S. K. Oh et al., 2013).

In this study, Raman spectroscopy was employed to monitor ethanol and glucose concentrations during the fermentation of the glucose obtained from the acid hydrolysis of brewer's spent grain (Lisci et al., 2022). The Partial Least Squares (PLS) algorithm was used to relate Raman spectra and concentration, while the variable importance in projection (VIP) was applied to reduce the number of regressor variables and improve model performance. The study was first carried out by considering standard solutions, and then samples obtained from the fermentation reactor were analyzed to assess the capability of the proposed approach to monitor a real bioprocess.

## 2. Raman spectroscopy

Raman spectroscopy is a valuable technique for the analysis of solid and liquid materials because of the high information content in the spectra and because it is a non-contact and non-destructive measurement methodology. Raman spectra processed with chemometric data analysis can lead to an accurate quantitative assessment of the chemical composition of complex mixtures.

The principle behind this analytical technique is here briefly summarized. Raman spectroscopy works by shining a monochromatic light source onto a sample and detecting the scattered light. When monochromatic light is incident on a molecule, most photons that undergo scattering are scattered elastically (Rayleigh scattering). A small amount of the scattered light shifts in energy from the laser frequency because of interactions between the incident electromagnetic waves and the vibrational energy levels of the molecules in the sample. Plotting the intensity of the shifted light against the frequency produces the Raman spectrum of the sample. The Raman scattered photons carry information about the chemical structure and identity of the material. Raman spectroscopy can also be employed for both qualitative and quantitative applications. Band areas are proportional to concentration, meaning Raman spectroscopy is amenable to straightforward quantitative analysis. The application of Raman spectroscopy is emerging in the biotechnological field since the production and purification of metabolites are essential for industrial production that covers chemicals, fuels or other ingredients (Lin, 2021). For further details, the reader can refer, e.g. to Notingher et al. (2017).

### 3. Materials and methods

#### 3.1 Samples

Standard solutions contained: ethanol (absolute for analysis, Carlo Erba reagents), and glucose (highly pure, Microbial Diagnostic).

Real samples were obtained from the fermentation process performed in duplicate in 2 L reactors (Applikon Biotechnology BV, Nieuwpoortweg, The Netherlands) operating anaerobically at 50 rpm,  $T=33.6\text{ }^{\circ}\text{C}$ ,  $\text{pH}=5.3$ , for 9 h. The working volume of the acid hydrolysate for each reactor was about 1.7 L. Before being inoculated, the fermentation medium was neutralized using 0.1 M NaOH until a pH of 5.3 was reached, and then 12.25% (v/v) of inoculum was added. During the experimental run, samples were taken at different times, filtered and stored at  $4^{\circ}\text{C}$  for determination of ethanol and sugars by HPLC and Raman spectroscopy analysis.

#### 3.2 Raman spectra

A Cora 5500 Raman spectrometer (Anton Paar), with a 532 nm laser wavelength, was employed in this study. The Raman equipment covers wavelengths between  $197.7\text{ cm}^{-1}$  and  $3500\text{ cm}^{-1}$ .

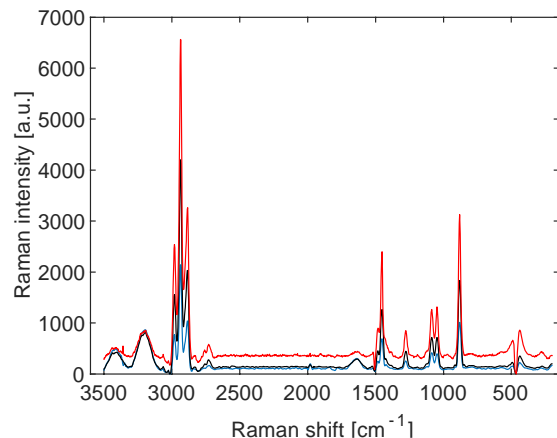


Figure 1. Spectra of samples with different ethanol concentrations: 5% blue line, 10% black line, 15% red line.

Table 1: Values of measurement parameters

Parameter	Values
Laser power	50 mW
Integration time	3000 ms
Average spectra	30

The wavelengths investigated here were included in the interval  $[764.6\text{ cm}^{-1} \div 2091\text{ cm}^{-1}]$ . Indeed, experimental noise significantly affects lower wavelengths, and no information about the species under investigation is appreciated at higher wavelengths. Since Raman spectroscopy is light-sensitive, each measurement was made in a dark environment. Measurement parameters are reported in Table 1. Baseline correction was carried out, and data were further post-processed using a noise-filtering technique.

Spectra obtained for some of the analyzed standard samples are reported in Figures 1-2. Dependence of the intensity from the concentration of ethanol can be visually appreciated in Figure 1, and the following characteristic peaks are clearly detected: i) stretching vibrations C-C at  $886\text{ cm}^{-1}$ ; ii) vibrations of the  $\text{CH}_3$  groups at  $1100\text{--}1116\text{ cm}^{-1}$ ; iii) torsion and rotational vibrations of the  $\text{CH}_2$  groups at  $1280\text{ cm}^{-1}$ ; iv) bending vibrations of  $\text{CH}_3$  and  $\text{CH}_2$  groups at  $1456\text{ cm}^{-1}$ ; v) stretching symmetric and asymmetric vibrations of the  $\text{CH}_3$  groups at  $3000\text{ cm}^{-1}$  approximately (Burikov et al., 2010). As many of the characteristic peaks of both species overlap when glucose is present, it is difficult to evaluate the data (Figure 2). This scenario has motivated the use of chemometric tools in order to discriminate the different contributions.

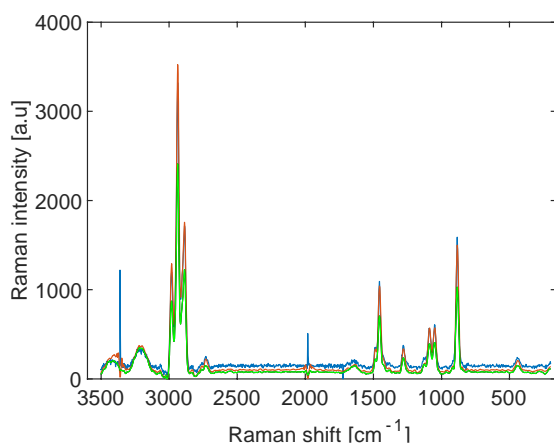


Figure 2. Spectra of samples with different glucose and ethanol concentration: 30 g/L of glucose and 15% v/v ethanol blue line, 15 g/L of glucose and 15% v/v ethanol red line, 22.5 g/L of glucose and 11.2% v/v ethanol green line.

#### 4. Statistical analysis

In Raman spectroscopy, sample complexity leads to spectral complexity, and in a fermentation system, it is often challenging to obtain unambiguous assignments of bands to specific analytes. Assignments are further complicated because many compounds of interest in solution are present at low concentrations, and their signal is overwhelmed by other high-concentration compounds, like water and the substrate. These difficulties imply that using statistical tools (chemometrics) to perform accurate and robust analyses is necessary.

The spectral data can be related to the concentration of the components in the samples by means of a PLS model (Geladi and Kowalski, 1986), which links the dependence of the multivariate data matrix  $\mathbf{X}$  (for the case at hand, the Raman spectra) to the composition of the same samples. The PLS model is obtained from a training set of  $N$  observations with  $L$   $X$ -variables, denoted by  $x_l$  ( $l = 1, \dots, L$ ) and  $M$   $Y$ -variables, denoted by  $y_m$  ( $m = 1, \dots, M$ ). The training data forms two matrices  $\mathbf{X}$  and  $\mathbf{Y}$  of dimension  $(N \times L)$  and  $(N \times M)$ . The PLS model is reported in Eq. 1, where the reference values  $\mathbf{Y}$  are the concentrations of ethanol and glucose

$$\begin{aligned} \mathbf{X} &= \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{TQ}^T + \mathbf{f} = \sum_{i=1}^A \mathbf{t}_i \mathbf{q}_i^T + \mathbf{F} \\ \mathbf{T} &= \mathbf{XW}(\mathbf{P}^T \mathbf{W})^{-1} \end{aligned} \quad (1)$$

In Eq. (1),  $\mathbf{E}$  ( $N \times L$ ) and  $\mathbf{F}$  ( $N \times M$ ) are error matrices containing the part of  $\mathbf{X}$  ( $N \times L$ ) and  $\mathbf{Y}$  ( $N \times M$ ), respectively, which the model does not explain,  $N$  and  $M$  are the number of samples (rows) and variables (columns), respectively. In the present study,  $N = 25$ , and  $M = 400$ . The vector  $\mathbf{t}_i$  is the  $i$ -th column vector that composes the score matrix  $\mathbf{T}$  ( $N \times A$ ),  $\mathbf{p}_i$  and  $\mathbf{q}_i$  are the loadings that compose the loading matrices  $\mathbf{P}$  ( $L \times A$ ) and  $\mathbf{Q}$  ( $M \times A$ ), where  $A$  is the number of latent variables chosen to explain the significant variance of the data. The matrix  $\mathbf{W}$  ( $L \times A$ ) is the weight matrix obtained by the PLS regression.

After PLS regression had been accomplished, the interest was focused on discriminating wavelengths in correspondence of which the absorption signal is the most influential on the changes in  $\mathbf{Y}$  (i.e., the concentrations) from those with no discrimination power. Many approaches have been proposed as variable selection methods. One of the most popular variable selection methods is the Variable Importance in Projection method (VIP), introduced by Wold et al. (1993), leading to the estimation of a so-called score for each regressor variable. VIP scores are useful in understanding the  $\mathbf{X}$  space predictor variables that best explain  $\mathbf{Y}$  variance.

For the  $j$ -th variable, the corresponding VIP score in a PLS model with  $A$  principal components can be calculated as:

$$VIP_j = \sqrt{L \frac{\sum_{a=1}^A (q_a^2 t_a^T t_a (\frac{w_{ja}}{\|w_a\|})^2)}{\sum_{a=1}^A q_a^2 t_a^T t_a}} \quad (2)$$

where  $t_a$  is the  $a$ -th column vector of the score matrix  $\mathbf{T}$ ,  $q_a$  is the  $a$ -th element of the regression coefficient vector  $q$  of  $\mathbf{T}$ ,  $w_a$  is the  $a$ -th column vector of the weighting matrix  $\mathbf{W}$ . Variables at which the VIP scores were above 1.0 were considered significant (Eriksson et al., 2006). The unity threshold value has been demonstrated to be a reasonable choice even when the  $\mathbf{X}$  variables are spectral data. The schematic representation of the proposed procedure is reported in Figure 3.

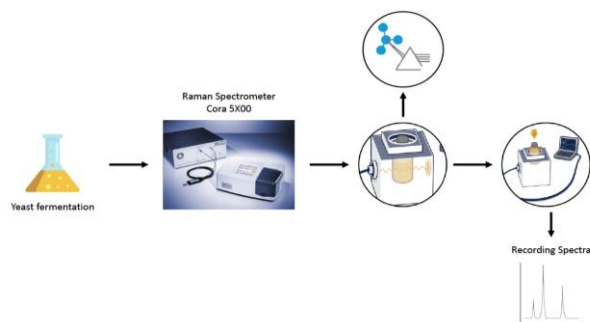


Figure 3. Schematic representation of the procedure for fermentation monitoring by Raman spectroscopy.

## 5. Results and discussion

PLS1D models have been developed for both ethanol and glucose by considering different aqueous solutions. It was found that 3 latent variables were satisfactory to describe the dependence on the ethanol, whereas, regarding the glucose dependence estimation, 4 latent variables were required. In Figure 4, VIP scores above a threshold of 1.0 are highlighted for ethanol and glucose.

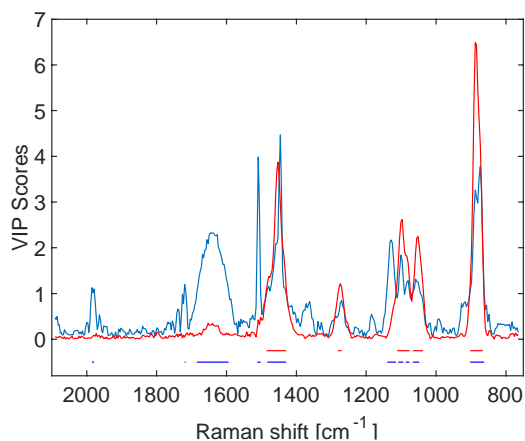


Figure 4. VIP scores, with the indication of the wavelengths with information on ethanol (red line) and glucose (blue line).

There is a significant overlap between the wavelengths informative for ethanol and glucose in most cases. VIPs result greater than one for both the species in wavelengths intervals located circa at  $880 \text{ cm}^{-1}$  (corresponding to the stretching vibrations of C-C), at  $1050 \text{ cm}^{-1}$  (corresponding to the  $\text{CH}_3$  vibrations), at  $1450 \text{ cm}^{-1}$  (related to the bending vibrations of  $\text{CH}_3$  and  $\text{CH}_2$ ) leading to some difficulties in discriminating the two components. In addition, the glucose shows VIPs statistically significant also at  $1100 \text{ cm}^{-1}$  (corresponding to COH vibrations, (see, e.g. Mathlouthi and Luu, 1980) and  $1600\text{-}1700 \text{ cm}^{-1}$ . This latter region should be related to water OH bending. Therefore, the statistical models relating spectra and concentrations for the two compounds have been constructed using only the interval of wavelengths suggested by the VIP scores.

The comparison between the ethanol concentration estimated with the PLS model with 3 latent variables and the experimental data is reported in Figure 5 (left panel). The results are quite satisfactory, as indicated by the coefficient of determination. Indeed,  $R^2$  is equal to 0.9927 for calibration and 0.9374 for validation. The glucose concentration values predicted by the PLS model with 4 latent variables are reported in Figure 5 (right panel) as a function of the measured ones. In this case, the coefficient of determination is equal to 0.9087 for calibration and 0.6237 for validation, indicating that glucose detection and quantification using the proposed approach are more challenging compared to alcohol.

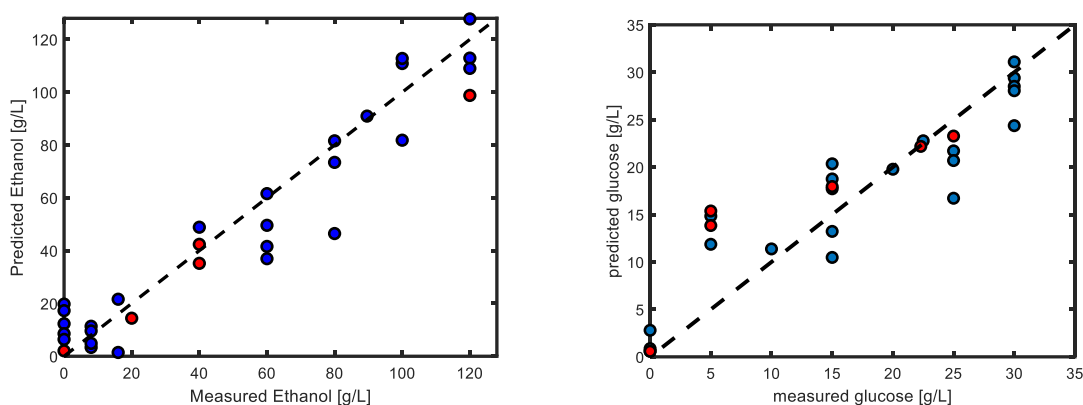


Figure 5. Predicted vs measured ethanol concentrations (left panel) and predicted vs measured glucose concentration (right panel) for the standard solutions. The red circles represent validation data.

PLS models were also obtained considering samples from the fermentation of hydrolyzed spent grain. The determination coefficients estimated in validation were equal to 0.7913 and 0.4783 for ethanol and glucose, respectively. Thus, the performance of the PLS models significantly deteriorated in this case. In fact, there is a possibility that the signal associated with suspended particles interfered with glucose quantification (Shih and Smith, 2009), indicating that samples filtering was not sufficient for obtaining acceptable results.

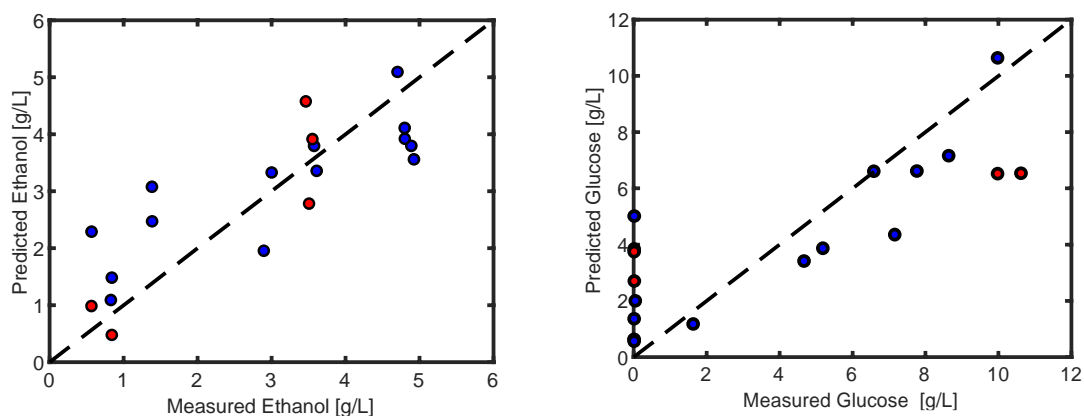


Figure 6. Predicted vs measured ethanol concentrations (left panel) and predicted vs measured glucose concentration (right panel) for the fermentation samples. The red circles represent validation data.

## 6. Conclusions

This study investigated and discussed the use of Raman spectroscopy as a monitoring tool for bioethanol production in fermentation bioreactors. The aim was to develop an automated sensor that could measure the composition of ethanol and glucose with minimal sample preparation and analysis time. Initially, the VIP variable selection method was employed to identify the most informative region of the spectra for the target compounds. Subsequently, PLS models were developed to establish a correlation between Raman spectra and the concentration of ethanol and glucose. The results obtained from analyzing standard solutions of ethanol and

glucose were satisfactory, as reported in previous investigations. However, analyzing fermentation solutions posed a considerable challenge due to the presence of suspended particles.

### Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778168.

### References

- Avila T.C., Poppi R.J., Lunardi I., Tizei P.A., Pereira G.A., 2012, Raman spectroscopy and chemometrics for on-line control of glucose fermentation by *Saccharomyces cerevisiae*, *Biotechnology Progress*, 28(6), 1598-1604.
- Burikov S., Dolenko T., Patsaeva S., Starokurov Y., Yuzhakov V., 2010, Raman and IR spectroscopy research on hydrogen bonding in water-ethanol systems, *Molecular Physics*, 108(18), 2427-2436.
- Claßen J., Aupert F., Reardon K.F., Solle D., Scheper T., 2017, Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application, *Analytical and Bioanalytical Chemistry*, 409, 651-666.
- Eriksson L., Kettaneh-Wold N., Trygg J., Wikström C., Wold S., 2006, Multi-and megavariable data analysis: Part I: basic principles and applications. In *Computer and Information Science* (Vol. 1). Umetrics Academy, Malmö, Sweden.
- 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.
- Geladi P., Kowalski B.R., 1986, Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185(C), 1-17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- Lin Y.K., Leong H.Y., Ling T.C., Lin D.Q., Yao S.J., 2021, Raman spectroscopy as process analytical tool in downstream processing of biotechnology, *Chinese Journal of Biotechnology*, 29(2), 204-211.
- Lisci S., Grosso M., Tronci S., 2020, A geometric observer-assisted approach to tailor state estimation in a bioreactor for ethanol production, *Processes* 8(4), 480.
- Lisci S., Tronci S., Grosso M., Karring H., Hajrizaj R., Errico M., 2022, Brewer's Spent Grain: its Value as Renewable Biomass and its Possible Applications, *Chemical Engineering Transactions*, 92, 259-264.
- Mathlouthi M., Luu D.V., 1980 Laser-Raman spectra of D-Glucose and Sucrose in aqueous solution, *Carbohydrate Research*, 81, 203-212.
- Mei C., Chen Y., Zhang H., Chen X., Liu, G., 2017, Development of a multi-model strategy based soft sensor using Gaussian process regression and principal component analysis in fermentation processes, *Chemical Engineering Transactions*, 61, 385-390.
- Notingher I., Sinjab F., Shipp D.W., 2017, Raman spectroscopy: Techniques and applications in the life sciences, *Advances in Optics and Photonics*, 9(2), 315-428.
- Oh S. K., Yoo S. J., Lee J.M., 2012, Predicting concentrations of a mixture in bioreactor for on-line monitoring using Raman spectroscopy, *IFAC Proceedings Volumes*, 45(15), 822-827.
- Shih C. J., Smith E.A., 2009, Determination of glucose and ethanol after enzymatic hydrolysis and fermentation of biomass using Raman spectroscopy, *Analytica Chimica Acta*, 653(2), 200-206.
- Wold S., Johansson A., Cochi M., 1993, PLS-partial least squares projections to latent structures, Chapter in: *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM Science Publishers, Germany, 523-550.