LWT - Food Science and Technology 155 (2022) 112951

Contents lists available at ScienceDirect

LWT

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

FT-NIR spectroscopy analysis for monitoring the microbial production of 2-phenylethanol using crude glycerol as carbon source

C. Quintelas [a,*,1], A. Braga [a,1], A. Cordeiro [a], E.C. Ferreira [a], I. Belo [a], R.N.M.J. Páscoa [b]

[a] CEB - Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
[b] LAQV/REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia, Universidade do Porto, Portugal

ARTICLE INFO

Keywords:
Biomass
Fragrance compounds
FT-NIR spectroscopy
Glycerol
PLS

ABSTRACT

The valorization of glycerol wastes through its conversion into high-added value products using biotechnological processes is very promising due to the high amount of glycerol generated from the bio-diesel industry. In this sense, it is important to develop rapid, reliable and, if possible, environmentally friendly analytical techniques for the monitoring of high-added value compounds. With this in mind, Fourier transform near infrared (FT-NIR) spectroscopy was explored for the assessment of biomass, glycerol, L-phenylalanine and 2-phenylethanol (2-PE) parameters during the biotechnological production of 2-PE. A total of 80 samples were collected and analyzed through FT-NIR spectroscopy. Partial least squares (PLS) regression method was selected for establishing quantification models. The obtained figures of merit, in terms of, coefficients of determination of prediction (R^2_P), range error-ratios (RER) and residual predictive deviation (RPD) yielded values higher than 0.85, 7.7 and 2.6, respectively, for all the parameters assessed. The obtained results reveal that all the developed PLS models can be considered as models with a good predictive accuracy and therefore attest the suitability of FT-NIR spectroscopy for the quantification of these parameters. Concluding, a new rapid, reliable, cost-effective, accurate, multi-parametric and environmentally friendly analytical technique exploring the potential of FT-NIR spectroscopy was successfully developed.

1. Introduction

The large production of glycerol, a waste from the biodiesel industry, is a threat to the economy and the environment. As biodiesel production grows, the environmental and sustainable valorization of crude glycerol becomes a fundamental task to reduce biodiesel production costs (Kaur et al., 2020). In this context, one possibility for glycerol valorization is its use as carbon source for microbial growth in bioprocesses (ex: for the production of high-added value products). The cosmetics and perfumery sectors are interested in fragrance compounds produced by biotechnological processes, since they are considered as natural compounds and accepted by the consumers (Berger, 2015). Worldwide, this industry generates a value exceeding 90 billion of euros, with a solid expected annual growth of 4–6 percent for next years (www.statista.com).

2-Phenylethanol (2-PE) is a higher aromatic alcohol with a characteristic aroma of roses. Flavor compounds are widely used in pharmaceutical, food, tobacco, perfumery, and cosmetic industries (Rodríguez-Romero et al., 2020). Moreover, 2-PE is a serious candidate for the next-generation of biofuels due to its high energy potential. In addition, it is also used for the production of derivatives such as styrene, phenylethyl acetate, and other valuable compounds (Liu et al., 2018). Flavors produced by biotechnological methods require the use of microorganisms for their production by *de novo* synthesis or through the biotransformation of appropriate precursors (Shaaban et al., 2016). The microorganisms that can be used to produce 2-PE include *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Pichia fermentans*, *Pichia anomala*, *Schizosaccharomyces pombe*, *Hansenula anomala* (Wang et al., 2019), *Saccharomyces cerevisiae*, *Escherichia coli* (Wang et al., 2019; Braga & Faria, 2020), *Wickerhamomyces anomalus*, *Candida glabrata*, *Candida utilis*, *Candida parapsilosis* (Rodriguez-Romero et al., 2020) and *Yarrowia lipolytica* (Gu et al., 2020).

Despite the advantages of using conventional carbon sources, such as glucose, lactose or sucrose, to allow high cell density cultures in biotransformation processes, the use of low-cost substrates, such as industrial by-products, makes the industrial biotechnological production of flavors economically advantageous and competitive.

* Corresponding author.
E-mail address: cquintelas@deb.uminho.pt (C. Quintelas).
[1] These authors contribute equally to this work.

https://doi.org/10.1016/j.lwt.2021.112951
Received 2 June 2021; Received in revised form 14 September 2021; Accepted 7 December 2021
Available online 11 December 2021
0023-6438/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Crude glycerol is the main by-product of biodiesel industry, where 10 kg of biodiesel produced generates 1 kg of glycerol. However, the generated glycerol is not pure and for that reason its use in traditional applications is limited (Johnson & Taconi, 2007). In addition, its purification is not a cost-effective process (Amaral et al., 2009; Wang et al., 2001). Thus, it becomes necessary to find alternative strategies for the recovery and valorization of this by-product in its un-purified form (Amaral et al., 2009; Çelik et al., 2008; Johnson & Taconi, 2007). The biotechnological conversion of crude glycerol by microbial fermentation into value-added products has been proposed as an interesting alternative (André et al., 2010; Koutinas et al., 2007). Glycerol can be assimilated as a carbon source by several microorganisms, being metabolized by the phosphorylation path to produce several compounds (Makri et al., 2010). The use of glycerol for *Y. lipolytica* growth and production of 2-PE from biotransformation of L-phenylalanine (L-Phe) has been reported by Celińska and Grajek (2013), with a titer of 1 g L^{-1} of this phenyl-alcohol. In yeasts, 2-PE can be produced through the Ehrlich pathway by L-Phe biotransformation.

The conventional methods applied for the quantification of glycerol and aromatic compounds are high performance liquid chromatography (HPLC) and gas chromatography (GC) (Pauls, 2011; Arellano et al., 2012; Hosseini et al., 2015). However, both methods present several disadvantages such as the impossibility of being used for real-time monitoring, they are reagent and time consuming and often it is not possible to obtain complete chromatographic separation, making it impossible to quantify the compounds. A faster and more direct method of analysis would produce more timely process information and better bioprocess control. Accurate monitoring of a bioprocess for the production of aromatic compounds, as 2-PE, is essential to achieve optimal quality and yield of this rose-like aromatic compound.

FT-NIR spectroscopy is an easy, rapid and nondestructive technique applied in many areas such as pharmaceutical, chemical, petrochemical, food and beverage, polymer and agriculture. Recently, developments in both chemometrics and instrumentation have resulted in rapid methods relating spectroscopic and multivariate chemical data that allowed to apply this technique for the monitoring of fermentative/biotransformation processes (Zimmerleiter et al., 2020). FT-NIR was applied for the monitoring of biomass, penicillin, and phenoxyacetic acid in a *Penicillium chrysogenum* fed-batch fermentation process (Zimmerleiter et al., 2020) and for the quantification of biomass, glucose, ethanol and glycerine concentrations in an alcoholic fermentation (Nascimento et al., 2017). The technique was also applied in the biodiesel quality assessment, through the determination of glycerol and alcohol (Pilar Dorado et al., 2011), in a rapid intracellular lipid quantification in oleaginous yeasts (Chmielarz et al., 2019) and, for the monitoring of 2-phenylethanol and L-Phe concentrations in culture medium of *Yarrowia lipolytica* (Quintelas et al., 2019).

In the present work the potential of FT-NIR spectroscopy was explored for monitoring a biotechnological process for the production of 2-PE measuring not only the precursor, but also the carbon source. In this process, crude glycerol from the biodiesel industry was used as carbon source to boost sustainability and economic viability of the production of biobased flavors. A rapid measurement of parameters as biomass, glycerol, L-phenylalanine and 2-PE is very important to ensure a high quality of the product as well as to optimize the quantity produced. A chemometric approach was essential with that regard, consisting of PLS regression to establish quantification models with the parameters abovementioned. The developed PLS models were optimized in terms of spectral regions and pre-processing techniques, regarding each parameter, and then the respective independent test set was projected to attest the robustness and accuracy of the PLS model.

2. Experimental

2.1. Microorganism, media and culture conditions

The microorganism used in this work was *Y. lipolytica* CH1/5 (isolated from cheeses), that was supplied by Dr. Nguyen and Dr. Szandra from Szent István University, Budapest (Nagy, 2014; Nagy et al., 2013).

For the bioconversion experiments, *Y. lipolytica* CH 1/5 cells were pre-cultivated at 27 °C and 200 rpm for 17 h in 500 mL baffled Erlenmeyer flasks containing 200 mL of YPG (10 g L^{-1} of yeast extract, 20 g L^{-1} of peptone and 20 g L^{-1} of glycerol) (Braga & Belo, 2015). Pre-cultures were centrifuged (8000 rpm, 10 min), washed with a NaCl solution (9 g L^{-1}) and used to inoculate the bioreactor (RALF PLUS SOLO, Bioengineering, Wald, Switzerland) with 1.1 L of the bioconversion of modified cultivation medium (based on Cui et al., 2011), containing per liter of deionized water: crude glycerol 40 g, KH_2PO_4 15 g, $MgSO_4.7H_2O$ 0.5 g, YNB without amino acids 0.2 g, thiamine 3 mg, supplemented with L-Phe 4 g. Samples of 5 mL were taken from the bioreactor three times per day, during 5 days. Bioconversion experiments were performed in duplicate.

The crude glycerol was supplied by Prio Energy - Prio Biocombustíveis, SA and has the following composition (w/w), according to the supplier: 82% glycerol, 11.8% water, 4.7% NaCl, less than 0.01% methanol, 0.5% organic matter (which is not glycerol) and less that 1% of other impurities.

A total of 80 samples were collected and analyzed by HPLC and FT-NIR. More details about the number of samples used for each chemometric model can be seen in Table 1.

2.2. Biomass, glycerol, 2-PE and L-Phe analysis

During the bioconversion experiments, samples were taken to determine biomass concentration, glycerol and L-Phe consumption and 2-PE production.

Cell concentration was determined by measuring the optical density (OD) at a wavelength of 600 nm. The OD values obtained were converted into dry cell weight (g L^{-1}) using a calibration curve previously obtained. The value of R^2 for the calibration was 0.999 and the limit of detection was 0.55 g L^{-1}.

The concentration of glycerol was determined by HPLC accordingly to Ferreira et al. (2016). The value of R^2 and limit of detection for the calibration were 0.9998 and 0.1 g L^{-1}, respectively.

The 2-PE and L-Phe concentrations were determined accordingly to Quintelas et al. (2019).

2.3. Near infrared scanning

FT-NIR spectra were recorded on a Fourier-transform near infrared spectrometer (FTLA 2000, ABB, Thermo Electron Corporation) equipped with an indium–gallium–arsenide (InGaAs) detector, from 14000 to 200 cm^{-1}, in transmittance mode using a flow cell with a 0.7 mm path length. For each sample, 64 scans were obtained with a spectral resolution of 8 cm^{-1} and then averaged. Samples were temperature equilibrated at 23 °C (approximately 3 min) within the spectrometer before scanning. The integration time was adjusted until the peaks at 1100–1200 nm for FT-NIR were close to 60,000 intensity units. Grams/AI software (Thermo Electron Corporation) was used for spectrometer configuration, control, and data acquisition. Distilled water was used as background.

2.4. Chemometrics analysis

Two chemometric tools, namely principal component analysis (PCA) (Jolliffe, 2002) and partial least squares (PLS) regression (Geladi & Kowalski, 1986), were used for the multivariate data analysis. The PCA tool was applied to check the presence of outliers and to detect common

Table 1
Statistics for the chemical parameters evaluated obtained by reference analysis.

Parameter	Calibration set				Validation set			
	N	Minimum	Maximum	Mean	N	Minimum	Maximum	Mean
Glycerol (g/L)	54	0.00632	48.1	19.40	23	0.0633	47.8	12.70
2-phenylethanol (g/L)	51	0.0255	0.428	0.18	22	0.0256	0.419	0.17
L-phenylalamine (g/L)	21	0.0332	4.91	2.57	9	0.134	4.58	2.78
Biomass (g/L)	56	0.446	29.2	14.00	24	0.487	27.9	11.20

tendencies. No outliers were detected, neither common tendencies. The PLS tool was applied for the quantification of glycerol, biomass, 2-PE and L-phenylalanine, using the PLS-1 algorithm. The spectra were previously mean centered before the application of PCA and PLS using Matlab R2014a version 8.3 (MathWorks, Natick, USA) using PLS Toolbox version 8.2.1 (Eigenvector Research Inc., Wenatchee, USA).

The sample set was initially divided into two data sets, one for calibration with 70% of the samples and the other set for validation with the rest of the samples (30%), using the Kennard-Stone algorithm. The division was made randomly but guaranteeing that the concentration of all the parameters in validation set were within the concentration values of the calibration set. Different spectral regions and pre-processing techniques were tested individually and in all possible combinations to verify which spectral region(s) and pre-processing techniques yielded the best results using only the calibration set. The spectral regions studied were the following: spectral region 1 from 5400 to 216 cm^{-1}, spectral region 2 from 6426 to 5404 cm^{-1}, spectral region 3 from 7217 to 6430 cm^{-1} and spectral region 4 from 13960 to 7221 cm^{-1}. The pre-processing techniques studied included standard normal variate (SNV) and Savitzky-Golay filter (tested with the first and second derivative, second polynomial order and different filter widths). The estimation of the optimal number of latent variables to prevent model overfitting was performed using leave-one-sample-out cross-validation method and the selection was based on the lowest root mean square error of cross-validation (RMSECV). The root mean square error of calibration (RMSEC), RMSECV, and root mean square error of prediction (RMSEP) were calculated according to the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(c_i - \hat{c}_i)^2}{n}} \quad (1)$$

where, n is the number of samples, c_i is the experimental measurement for sample i and $\hat{c}_i$ is the corresponding value obtained for calibration (RMSEC), cross-validation (RMSECV) and prediction set (RMSEP).

The best PLS calibration model, for each parameter, was selected based on the lowest RMSEC and RMSECV and on the highest R^2_C and R^2_{CV}.

Then, after this optimization, the independent validation set was projected onto the PLS calibration models to assess their accuracy. The RMSEP, coefficient of determination of prediction (R^2_P), range error-ratio (RER) and residual predictive deviation (RPD) parameters were used to estimate the accuracy of the developed models. The RER and RPD parameters were calculated using the following formulas:

$$RER = \frac{(Y_{max} - Y_{min})}{RMSEP} \quad (2)$$

$$RPD = \frac{Standard\ deviation\ of\ reference\ data}{SEP} \quad (3)$$

In eq. (3), *SEP* stands for standard error of prediction (SEP) which was calculated using the following formula:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(c_i - \hat{c}_i - bias)^2}{n-1}} \quad (4)$$

where, c_i is the experimental results for sample i, $\hat{c}_i$ is the value obtained with the validation and n is the total number of samples used in the validation.

3. Results and discussion

3.1. Statistics for the chemical parameters evaluated

The PCA analysis did not reveal the formation of clusters or the presence of outliers (through the analysis of on Hotelling's T^2 and squared residuals statistics). Thus, all the available data were considered for further analysis and divided into two groups, one for calibration (70%) and the other for validation (30%). The minimum, maximum, and mean values of the calibration and validation sets for each parameter are shown in Table 1.

The number of samples available for each parameter evaluated is unequal because some samples yielded concentration values below the detection limit of the respective reference procedures used and for that reason were excluded. As can be seen, the values of the validation set are within the values of the calibration set.

3.2. Optimization of the PLS models

The optimization of the PLS calibration models was performed using only the calibration set. The models were optimized in terms of spectral regions and pre-processing techniques were used. As referred in section 2.4 Chemometric analysis, the spectra were divided in 4 spectral regions. These regions were tested individually and in all possible combinations (a total of 15 possible combinations) using different pre-processing techniques. The obtained results for the calibration models considering each chemical parameter individually are shown in Table 2. Note that only the application of Savitzky-Golay filter with 15 points of filter width and also the best spectral regions (spectral regions R2 and R4) are shown as the obtained results with other filter widths and spectral regions did not yield better results.

The obtained results (in terms of RMSEC, RMSECV, R^2_C and R^2_{CV}) helped to reveal the best spectral region and pre-processing technique for all the chemical parameters assessed. In terms of spectral regions, the best results considering the parameters abovementioned were obtained through the combination of spectral regions 2 and 4, which makes sense since these were the most informative spectral regions. Regarding the best pre-processing technique and considering all the parameters, the application of Savitzky-Golay filter with 15 points of filter width, second polynomial order and first derivative yielded the best results.

3.3. Validation of the PLS models

After the optimization of the PLS models using only the calibration set, the validation set was projected to test the accuracy of the developed models. Table 3 and Fig. 1 presents the obtained results.

As can be seen in Table 3, the obtained R^2_P was higher than 0.85 for all the compounds assessed, and for the quantification of glycerol and biomass it was even higher than 0.95. Following the same tendency, the RER parameter was higher than 7.7 for all the parameters and close to 15 for glycerol and biomass. Therefore, these PLS models can be considered very good models for research quantification (Rambo et al., 2013). Regarding the RPD values, all the developed models yielded values

Table 2
PLS calibration models results.

Glycerol																		
Pre-processing technique	Without			SNV			SG(15,2,1)+SNV			SG(15,2,2)+SNV			SG(15,2,1)			SG(15,2,2)		
Spectral regions	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4
RMSEC	1.72	2.08	1.84	3.94	3.55	3.14	4.72	4.09	3.43	2.14	5.07	2.15	1.73	1.70	1.27	2.30	4.79	1.40
RMSECV	3.02	3.53	2.95	6.38	6.77	4.43	6.86	6.31	5.78	3.50	10.81	4.03	2.93	3.59	2.74	3.00	8.19	3.05
R^2_C	0.95	0.94	0.94	0.95	0.96	0.97	0.93	0.95	0.96	0.94	0.92	0.94	0.99	0.99	0.99	0.96	0.91	0.97
R^2_{CV}	0.93	0.91	0.92	0.88	0.86	0.94	0.86	0.87	0.90	0.91	0.64	0.91	0.97	0.96	0.98	0.95	0.78	0.95
2-phenyethanol																		
Pre-processing technique	Without			SNV			SG(15,2,1)+SNV			SG(15,2,2)+SNV			SG(15,2,1)			SG(15,2,2)		
Spectral regions	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4
RMSEC	0.04	0.03	0.03	0.05	0.05	0.04	0.05	0.04	0.05	0.03	0.08	0.04	0.04	0.04	0.03	0.03	0.04	0.03
RMSECV	0.05	0.05	0.05	0.06	0.07	0.05	0.06	0.05	0.05	0.05	0.09	0.06	0.05	0.05	0.05	0.05	0.07	0.07
R^2_C	0.92	0.94	0.93	0.87	0.85	0.89	0.84	0.92	0.87	0.93	0.61	0.94	0.94	0.90	0.94	0.93	0.93	0.93
R^2_{CV}	0.86	0.82	0.85	0.77	0.71	0.83	0.79	0.84	0.83	0.84	0.50	0.81	0.87	0.84	0.87	0.84	0.70	0.85
L-phenylalanine																		
Pre-processing technique	Without			SNV			SG(15,2,1)+SNV			SG(15,2,2)+SNV			SG(15,2,1)			SG(15,2,2)		
Spectral regions	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4
RMSEC	0.85	0.76	0.54	0.70	0.72	0.78	0.75	0.13	0.71	0.23	0.12	0.20	0.61	0.55	0.57	0.24	0.01	0.05
RMSECV	0.98	1.17	1.24	1.06	1.19	0.91	1.42	1.58	1.35	1.97	2.55	1.87	0.98	1.23	1.16	1.90	1.66	1.77
R^2_C	0.79	0.82	0.89	0.83	0.82	0.80	0.89	0.89	0.88	0.88	0.89	0.88	0.86	0.89	0.88	0.88	0.89	0.89
R^2_{CV}	0.66	0.55	0.49	0.60	0.53	0.72	0.78	0.76	0.80	0.59	0.13	0.57	0.86	0.85	0.86	0.70	0.11	0.57
Biomass																		
Pre-processing technique	Without			SNV			SG(15,2,1)+SNV			SG(15,2,2)+SNV			SG(15,2,1)			SG(15,2,2)		
Spectral regions	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4	2	4	2 + 4
RMSEC	1.83	2.05	1.97	2.73	2.67	2.03	2.51	2.62	2.26	2.34	1.94	1.87	1.83	1.72	1.82	1.79	1.89	1.89
RMSECV	2.35	2.48	2.47	3.24	3.45	2.98	4.17	4.23	3.64	2.80	3.75	2.87	2.14	2.33	2.17	2.32	3.56	2.51
R^2_C	0.94	0.91	0.91	0.89	0.90	0.91	0.91	0.90	0.91	0.91	0.94	0.93	0.95	0.99	0.95	0.92	0.94	0.92
R^2_{CV}	0.89	0.89	0.87	0.85	0.83	0.87	0.77	0.76	0.82	0.89	0.80	0.88	0.91	0.92	0.93	0.90	0.81	0.89

Legend: SNV – standard normal variate; SG (x,y,z) – Savitzky-Golay (filter width points, polynomial order, derivative); RMSEC – root mean square error of calibration; RMSECV – root mean square error of cross-validation; R^2c – coefficient of determination for the calibration; R^2cv – coefficient of determination for the cross-validation.

Table 3
PLS results of the quantification of glycerol, 2-phenylethanol, L-phenylalamine and biomass when using the validation set.

	Glycerol	2-phenylethanol	L-phenylalanine	Biomass
LV	4	5	5	5
RMSEC (g/L)	1.27	0.0301	0.570	1.82
RMSECV (g/L)	2.74	0.0463	1.16	2.17
RMSEP (g/L)	2.72	0.0473	0.575	1.88
R^2_C	0.994	0.944	0.877	0.951
R^2_{CV}	0.975	0.870	0.860	0.931
R^2_P	0.969	0.850	0.876	0.958
RER	17.6	8.4	7.7	14.6
RPD	5.7	3.0	2.6	4.3

Legend: LV – latent variables; RMSEC – root mean square error of calibration; RMSECV – root mean square error of cross-validation; RMSEP – root mean square error of prediction; R^2c – coefficient of determination for the calibration; R^2cv – coefficient of determination for the cross-validation; R^2_P – coefficient of determination for the prediction; RER – range error-ratio; RPD – residual predictive deviation.

higher than 2.6 which indicates that all the models can be considered with a good predictive accuracy and the models with an RPD higher than 3.0 can be considered with an excellent accuracy according to (Nicolaï et al., 2007).

From all the PLS models developed, the quantification of L-phenylalanine was the one that produced the worst results which can be attributed to the low number of samples. Therefore, further studies including a higher number of samples are needed to attest the reliability of the model. Nonetheless, the developed PLS model for L-phenylalanine quantification can be considered with good predictive accuracy as the obtained RPD value was higher than 2 (Nicolaï et al., 2007).

Globally, the obtained results point to the fact that is possible to accurately quantify these parameters in biotechnological processes, namely fermentations, using the NIR spectra.

3.4. Analysis of regression coefficient vectors

With the objective of analyzing the most important wavenumbers for each of the developed PLS models, the regression coefficient vectors were depicted (Fig. 2).

The analysis of the regression coefficient vectors revealed that the most important wavenumbers for each of PLS model were very similar. For glycerol, the wavenumbers located within 7400–7260 and 6400 - 5400 cm^{-1} were the most important, being the wavenumbers around 5900 cm^{-1} the most important ones. The wavenumbers within 7400–7260 cm^{-1} can be attributed to C-H stretching on combination bands region while the wavenumbers within 6400–5400 cm^{-1} can be associated with C-H stretching on the first overtone region (Shenk et al., 2008). The main spectral differences found between samples of biodiesel with and without contamination of glycerol were located at 5814 and 5714 cm^{-1} (Dorado et al., 2011) which reinforces these findings. Moreover, the O-H stretching of alcohols can be found around 6300 cm^{-1} (Sieslet et al., 2008).

For 2-phenylethanol and L-phenylalanine, the more relevant wavenumbers were located within 6400–5400 cm^{-1}. Again, these spectral regions can be connected with C-H stretching on the first overtone region (Shenk et al., 2008) and the O-H stretching of alcohols around 6300 cm^{-1} (Sieslet et al., 2008).

In relation to biomass, the wavenumbers within 7370–7260 and 6400 - 5400 cm^{-1} were the most important. Theses wavenumbers can be

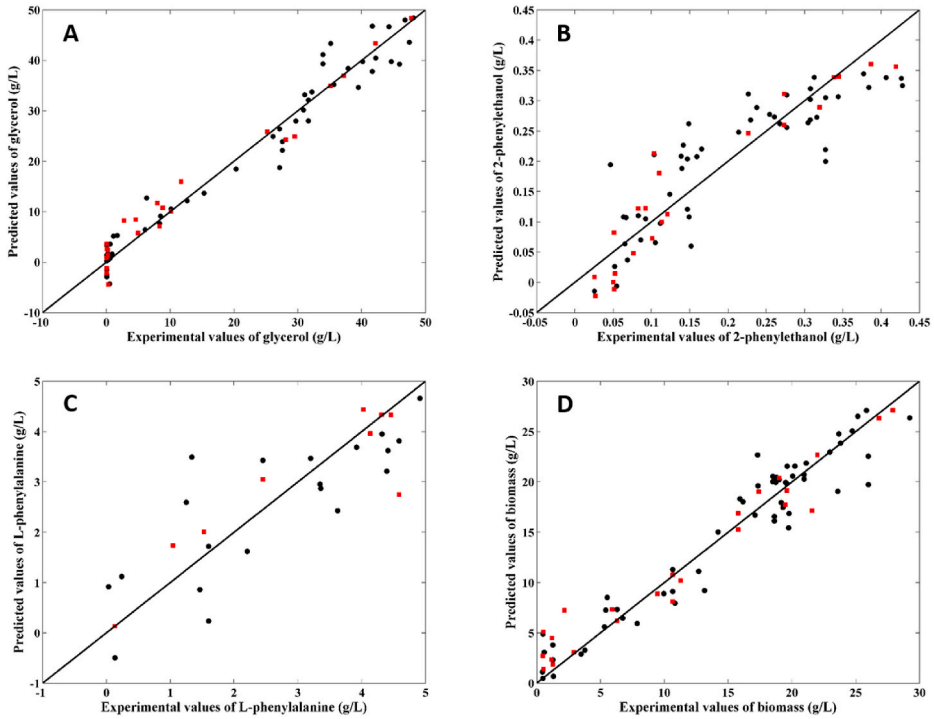


Fig. 1. Experimental values versus the cross-validation (●) and validation (■) values obtained for (A) glycerol, (B) 2-phenylethanol, (C) L-phenylalanine and (D) biomass.

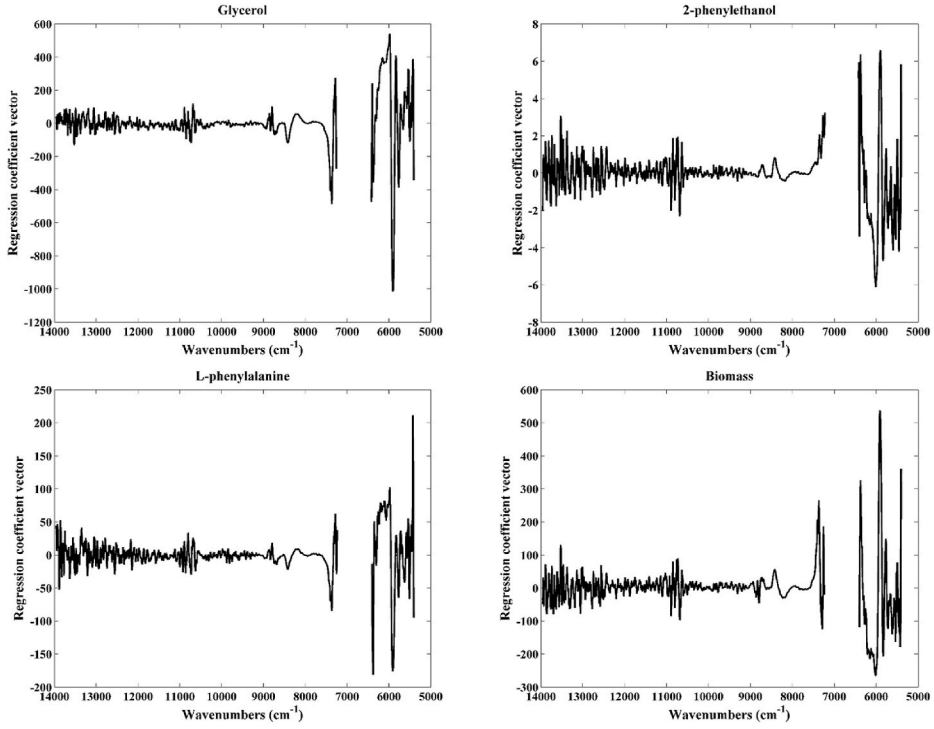


Fig. 2. Regression coefficients vectors of the PLS models built for glycerol, 2-phenylethanol, L-phenylalanine and biomass.

connected with C-H stretching on combination bands regions and first overtone region, respectively (Shenk et al., 2008).

In our opinion, the most important wavenumbers found for each PLS model make sense as the chemical bonds that were previously referred, are present in each evaluated parameter.

3.5. Remarks and perspective

Several reports described the use of FT-NIR spectroscopy for monitoring different fermentation processes. To improve the yield of an industrial fermentation, Wang et al. (2017) applied a method based on Fourier-transform near-infrared spectroscopy (FT-NIR) to predict the growth of the yeast. These authors developed a model based on gaussian

mixture regression (GMR) and obtained a coefficient of determination R_p^2 of 0.98 for yeast quantification, that is similar to the one obtained in the present report. Grassi et al. (2014) tested the application FT-NIR for the quantification of biomass (yeasts) in a beer fermentation using PLS regression and locally weighted regression. The R_p^2 ranged between 0.829 and 0.852. The quantitative monitoring of substrate consumption, namely glucose, was also successfully accomplished through FT-NIR (Nascimento et al. (2017) and Tamburini et al. (2014)). Morita et al. (2011) analyzed the glucose and xylose consumption in a fermentation process aiming to produce ethanol, glycerol, and xylitol. The authors developed robust PLS regression models for all the compounds, with a high prediction accuracy (standard error of prediction values in the range 0.21–1.49 g L^{-1}). The quantification of glycerol by FT-NIR spectroscopy was mainly applied to the production/analysis of biodiesel (Gelinski et al., 2018; Furtado et al., 2021). Dorado et al. (2011) developed a methodology based in FT-NIR spectroscopy for the quantification of methanol and glycerol and achieved a R^2_{CV} of 0.99 and 0.81, for methanol and glycerol, respectively. Taking into consideration that the current study obtained an R_P^2 value of 0.97 for glycerol determination, – higher than that obtained by the previous authors – this reinforces the higher performance of the developed methodology. The quantification of 2-PE and L-Phe by FT-NIR, in fermentative processes, is not usual, and this technique is mainly applied to the quantification of 2-PE in wines and for the quantification of amino acids, as L- Phe, in food products. Studies developed by Lorenzo et al. (2009) and Ye et al. (2016) tested the quantification of 2-PE in wines, red wines and apple wines, respectively. The first authors obtained a R_P^2 of 0.36 and the second authors obtained a R_P^2 of 0.84 for 2-PE estimation. Again, the developed work shows a higher performance over the cited methodologies. In fact, an R_P^2 value of 0.88 was obtained, which compares quite favorably with the 0.36 value of Lorenzo et al. (2009) and 0.84 value of Ye et al. (2016). The quantification of L-Phe using FT-NIR spectroscopy was tested by Escuredo et al. (2014) that analyzed the presence of this amino acid plus other eleven amino acids in quinoa and by Shen et al. (2010) that evaluated the amino acids content, including L- Phe, in Chinese rice wine. The first authors obtained a coefficient of determination (R^2) of 0.78, for L- Phe and the second authors found an R^2 of 0.89 for L-Phe determination. Taking into consideration that the current study obtained an R_P^2 value of 0.88 for L-Phe determination, which is higher than the obtained by Escudero et al. and similar to the obtained by Shen et al., this attests again the high performance of the developed methodology.

It is important to highlight that the obtained concentrations for the production of 2-PE using glycerol as carbon source are similar with the ones obtained with glucose, which reinforces the possibility of using this waste, crude glycerol, to produce value added products (Braga et al., 2018; Ferreira et al., 2016, 2018, pp. 41–42). The transformation of waste or by-products from different industries into add-value products for a circular economy and the fast and reliable determination of the parameters associated to the process is extremely important. It should be emphasized that this methodology presents several advantages over classical analytical techniques, offering a practical alternative to time-consuming methods such as liquid or gas chromatographic techniques.

4. Conclusions

The monitoring of biotechnological processes demands the use of rapid, reliable and, if possible, environmentally friendly analytical techniques. In this sense, this work describes a rapid, reliable, cost-effective, accurate, multi-parametric and environmentally friendly analytical technique exploring the potential of FT-NIR spectroscopy for the assessment of biomass, glycerol, L-phenyalanine and 2-PE during a biotechnological process for the production of 2-PE.

The determination of these parameters was achieved though PLS regression. The PLS calibration models were optimized in terms of spectral regions and pre-processing techniques. The combination of spectral regions 2 (6426–5404 cm^{-1}) and 4 (13960–7221 cm^{-1}) with the spectra pre-processed using Savitzky-Golay filter (with 15 points of filter width, second polynomial order and first derivative) yielded the best results. After the optimization, the models were validated using an independent test set and the obtained results in terms of R^2_P, RER and RPD were higher than 0.85, 7.7 and 2.6, respectively, for all the parameters assessed. These results reveal that all the developed PLS models can be considered as models with a good predictive accuracy and as so attest the suitability of FT-NIR spectroscopy for the quantification of these parameters. The most important wavenumbers, analyzed through the respective regression coefficient vectors, pointed to chemical bonds that are present in each of the evaluated parameters.

CRediT authorship contribution statement

C. Quintelas: Conceptualization, Methodology, Investigation, Writing – original draft. **A. Braga:** Conceptualization, Methodology, Investigation. **A. Cordeiro:** Investigation. **E.C. Ferreira:** Writing – review & editing, Funding acquisition. **I. Belo:** Writing – review & editing, Funding acquisition. **R.N.M.J. Páscoa:** Formal analysis, Writing – original draft.

Declaration of competing interest

There is no conflict of interest.

Acknowledgments

The authors thank the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic fundings of UIDB/BIO/04469/2020 and UIDB/QUI/50006/2020 units. Ricardo N.M.J. Páscoa and Cristina Quintelas thanks FCT for funding through program DL 57/2016 – Norma transitória. The authors also want to express their gratitude to Prio Energy – Prio Biocombustíveis, SA for providing the glycerol used in this experimental work.

References

Amaral, P., Ferreira, T., Fontes, G., & Coelho, M. (2009). Glycerol valorization: New biotechnological routes. *Food and Bioproducts Processing, 87*, 179–186. https://doi.org/10.1016/j.fbp.2009.03.008

André, A., Diamantopoulou, P., Philippoussis, A., Sarris, D., Komaitis, M., & Papanikolaou, S. (2010). Biotechnological conversions of bio-diesel derived waste glycerol into added-value compounds by higher fungi: Production of biomass, single cell oil and oxalic acid. *Industrial Crops and Products, 31*, 407–416. https://doi.org/10.1016/j.indcrop.2009.12.011

Arellano, M., Gschaedler, A., & Alcazar, M. (2012). Major volatile compounds analysis produced from mezcal fermentation using gas chromatography equipped headspace (GC–HS). In D. B. Salih (Ed.), *Gas chromatography in plant science, wine Technology, toxicology and some specific applications*. InTech. https://doi.org/10.5772/33415.

Berger, R. (2015). Biotechnology as a source of natural volatile flavours. *Current Opinion in Food Science, 1*, 38–43. https://doi.org/10.1016/j.cofs.2014.09.003

Braga, A., & Belo, I. (2015). Production of gamma-decalactone by Yarrowia lipolytica: Insights into experimental conditions and operating mode optimization. *Journal of Chemical Technology and Biotechnology, 90*, 559–565. https://doi.org/10.1002/jctb.4349

Braga, A., & Faria, N. (2020). Bioprocess optimization for the production of aromatic compounds with metabolically engineered hosts: Recent developments and future challenges. *Frontiers in Bioengineering and Biotechnology, 8*(1–18), 96. https://doi.org/10.3389/fbioe.2020.00096

Braga, A., Oliveira, A., Freitas, B., Szandra, N. E., Nguyen, Q. D., & Belo, I. (2018). Yarrowia lipolytica as a potential producer of 2-phenylethanol from L-phenylalanine biotransformation. CHEMPOR 2018. In *13th international chemical and biological engineering conference (book of extended abstracts). Aveiro, Portugal, oct 2-4*. http://hdl.handle.net/1822/56573.

Çelik, E., Ozbay, N., Oktar, N., & Çalik, P. (2008). Use of biodiesel byproduct crude glycerol as the carbon source for fermentation processes by recombinant Pichia pastoris. *Industrial & Engineering Chemistry Research, 47*, 2985–2990. https://doi.org/10.1021/ie071613o

Celińska, E., & Grajek, W. (2013). A novel multigene expression construct for modification of glycerol metabolism in Yarrowia lipolytica. *Microbial Cell Factories, 12*, 1–16. https://doi.org/10.1186/1475-2859-12-102

Chmielarz, M., Sampels, S., Blomqvist, J., Brandenburg, J., Wende, F., Sandgren, M., & Passoth, V. (2019). FT-NIR: A tool for rapid intracellular lipid quantification in oleaginous yeasts. *Biotechnology for Biofuels, 12*(1–9), 169. https://doi.org/10.1186/s13068-019-1513-9

Cui, Z., Yang, X., Shen, Q., Wang, K., & Zhu, T. (2011). Optimisation of biotransformation conditions for production of 2-phenylethanol by a Saccharomyces cerevisiae CWY132 mutant. *Natural Product Research, 25*, 754–759. https://doi.org/10.1080/14786419.2010.529441

* Dorado, M. P., Pinzi, S., de Haro, A., Font, R., & Garcia-Olmo, J. (2011). Visible and NIR Spectroscopy to assess biodiesel quality: Determination of alcohol and glycerol traces. *Fuel, 90*, 2321–2325. https://doi.org/10.1016/j.fuel.2011.02.015.

Escuredo, O., Martín, M. I. G., Moncada, G. W., Fischer, S., & Hierro, J. M. H. (2014). Amino acid profile of the quinoa (Chenopodium quinoa Willd.) using near infrared spectroscopy and chemometric techniques. *Journal of Cereal Science, 60*, 67–74. https://doi.org/10.1016/j.jcs.2014.01.016

Ferreira, P., Lopes, M., Mota, M., & Belo, I. (2016). Oxygen transfer rate and pH as major operating parameters of citric acid production from glycerol by Yarrowia lipolytica W29 and CBS 2073. *Chemical Papers, 70*, 869–876. https://doi.org/10.1515/chempap-2016-0024

Ferreira, P., Ribeiro, M., Machado, A. R., Nagy, E., Nguyen, & D. Q., & Belo, I. (2018). Production of erythritol by Yarrowia species from crude glycerol. CHEMPOR 2018. In *13th international chemical and biological engineering conference (book of extended abstracts). No. O-BB01, aveiro, Portugal, oct 2-4*. http://hdl.handle.net/1822/56846.

Furtado, W. L., Corgozinho, C. N. C., Tauler, R., & Sena, M. M. (2021). Monitoring biodiesel and its intermediates in transesterification reactions with multivariate curve resolution alternating least squares calibration models. *Fuel, 283*(1–11), Article 119275. https://doi.org/10.1016/j.fuel.2020.119275

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9

Gelinski, E. K., Hamerski, F., Corazza, M. L., & Santos, A. F. (2018). Biodiesel synthesis monitoring using near infrared spectroscopy. *The Open Chemical Engineering Journal, 12*, 95–110. https://doi.org/10.2174/1874123101812010095

Grassi, S., Amigo, J. M., Lyndgaard, C. B., Foschino, R., & Casiraghi, E. (2014). Beer fermentation: Monitoring of process parameters by FT-NIR and multivariate data analysis. *Food Chemistry, 155*, 279–286. https://doi.org/10.1016/j.foodchem.2014.01.060

Hosseini, R., Naderi, F., & Nasrollahi, S. A. (2015). Determination of phenylethyl alcohol by reversed phase high-performance liquid chromatography (RPHPLC) in Budesonide nasal spray. *African Journal of Pure and Applied Chemistry, 9*, 81–90. https://doi.org/10.5897/AJPAC2015.0619. https://www.statista.com/topics/1008/cosmetics-industry/

Johnson, D., & Taconi, K. (2007). The glycerin gut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environmental Progress, 26*, 338–348. https://doi.org/10.1002/ep.10225

Jolliffe, I. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag. https://doi.org/10.1007/b98835

Kaur, J., Sarma, A. K., Jha, M. K., & Gera, P. (2020). Valorisation of crude glycerol to value-added products: Perspectives of process technology, economics and environmental issues. *Biotechnology Reports, 27*, Article e00487. https://doi.org/10.1016/j.btre.2020.e00487

Koutinas, A., Wang, R.-H., & Webb, C. (2007). The biochemurgist - bioconversion of agricultural raw materials for chemical production. *Biofuels Bioproducts and Biorefining, 1*, 24–38. https://doi.org/10.1002/bbb.6

Liu, C., Zhang, K., Cao, W., Zhang, G., Chen, G., Yang, H., Wang, Q., Liu, H., Xian, M., & Zhang, H. (2018). Genome mining of 2-phenylethanol biosynthetic genes from *Enterobacter* sp. CGMCC 5087 and heterologous overproduction in *Escherichia coli*. *Biotechnology for Biofuels, 11*(1–15), 305. https://doi.org/10.1186/s13068-018-1297-3

* Lorenzo, C., Garde-Cerdán, T., Pedroza, M. A., Alonso, G. L., & Salinas, M. R. (2009). Determination of fermentative volatile compounds in aged red wines by near infrared spectroscopy. *Food Research International, 42*, 1281–1286. https://doi.org/10.1016/j.foodres.2009.03.021.

Makri, A., Fakas, S., & Aggelis, G. (2010). Metabolic activities of biotechnological interest in Yarrowia lipolytica grown on glycerol in repeated batch cultures. *Bioresource Technology, 101*, 2351–2358. https://doi.org/10.1016/j.biortech.2009.11.024

Morita, H., Hasunuma, T., Vassileva, M., Tsenkova, R., & Kondo, A. (2011). Near infrared spectroscopy as high-throughput technology for screening of xylose-fermenting recombinant Saccharomyces cerevisiae strains. *Analytical Chemistry, 83*, 4023–4029. https://doi.org/10.1021/ac103128p

Nagy, E. (2014). Isolation and diversity of food spoilage Yarrowia yeast strains from meat. *Acta Alimentaria, 43*, 101–106. https://doi.org/10.1556/AAlim.43.2014.Suppl.15

Nagy, E., Niss, M., Dlauchy, D., Arneborg, N., Nielsen, D., & Péter, G. (2013). Yarrowia divulgata f.a., sp. nov., a yeast species from animal-related and marine sources. *International Journal of Systematic and Evolutionary Microbiology, 63*, 4818–4823. https://doi.org/10.1099/ijs.0.057208-0

Nascimento, R. J. A., Macedo, G. R., Santos, E. S., & Oliveira, J. A. (2017). Real time and in situ near-infrared spectroscopy (NIRS) for quantitative monitoring of biomass, glucose, ethanol and glycerine concentrations in an alcoholic fermentation. *Brazilian Journal of Chemical Engineering, 34*, 459–468. https://doi.org/10.1590/0104-6632.20170342s20150347

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*, 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Pauls, R. E. (2011). A review of chromatographic characterization techniques for biodiesel and biodiesel blends. *Journal of Chromatographic Science, 49*, 384–396. https://doi.org/10.1093/chromsci/49.5.384

Quintelas, C., Braga, A., Mesquita, D., Amaral, A. L., Ferreira, E. C., & Belo, I. (2019). NIR spectroscopy applied to the determination of 2-phenylethanol and L-phenylalanine concentrations in culture medium of Yarrowia lipolytica. *Journal of Chemical Technology and Biotechnology, 94*, 812–818. https://doi.org/10.1002/jctb.5827

Rambo, M. K. D., Amorim, E. P., & Ferreira, M. M. C. (2013). Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. *Analytica Chimica Acta, 775*, 41–49. https://doi.org/10.1016/j.aca.2013.03.015

Rodríguez-Romero, J., Aceves-Lara, C., Silva, C., Gschaedler, A., Amaya-Delgado, L., & Arrizon, J. (2020). 2-Phenylethanol and 2-phenylethylacetate production by nonconventional yeasts using tequila vinasses as a substrate. *Biotechnology Reports, 25*, Article e00420. https://doi.org/10.1016/j.btre.2020.e00420

Shaaban, H., Mahmoud, K., Amin, A. A., & El Banna, H. A. (2016). Application of biotechnology to the production of natural flavor and fragrance chemicals. *Research Journal of Pharmaceutical, Biological and Chemical Sciences, 7*, 2670–2717.

* Shenk, J. S., Workman, J. J., & Westerhaus, M. O. (2008). Application of NIR spectroscopy to agricultural products. In D. A. Burns, & E. W. Ciurczak (Eds.), *Handbook of near-infrared analysis* (3rd ed.). CRC Press, Taylor and Francis Group, ISBN 9780849373930.

This reference is a key reference because it allowed us to define which wavelengths are most appropriate for the development of PLS models.

Shen, F., Niu, X., Yang, D., Ying, Y., Li, B., Zhu, G., & Wu, J. (2010). Determination of amino acids in Chinese rice wine by Fourier transform near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 58*, 9809–9816. https://doi.org/10.1021/jf1017912

Tamburini, E., Marchetti, M. G., & Pedrini, P. (2014). Monitoring key parameters in bioprocesses using near-infrared technology. *Sensors, 14*, 18941–18959. https://doi.org/10.3390/s141018941

Wang, W., Jiang, H., Liu, G., Chen, Q., Mei, C., Li, K., & Huang, Y. (2017). Quantitative analysis of yeast growth process based on FT-NIR spectroscopy integrated with Gaussian mixture regression. *RSC Advances, 7*, 24988–24994. https://doi.org/10.1039/C7RA02774E

Wang, Y., Zhang, H., Lu, X., Zong, H., & Zhuge, B. (2019). Advances in 2-phenylethanol production from engineered microorganisms. *Biotechnology Advances, 37*, 403–409. https://doi.org/10.1016/j.biotechadv.2019.02.005

Wang, Z.-X., Zhuge, J., Fang, H., & Prior, B. (2001). Glycerol production by microbial fermentation: A review. *Biotechnology Advances, 19*, 201–223. https://doi.org/10.1016/j.biotechadv.2019.02.005

* Ye, M., Gao, Z., Li, Z., Yuan, Y., & Yue, T. (2016). Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. *Food Chemistry, 190*, 701–708. https://doi.org/10.1016/j.foodchem.2015.05.112.

* Zimmerleiter, R., Kager, J., Nikzad-Langerodi, R., Berezhinskiy, V., Westad, F., Herwig, C., & Brandstetter, M. (2020). Probeless non-invasive near-infrared spectroscopic bioprocess monitoring using microspectrometer technology. *Analytical and Bioanalytical Chemistry, 412*, 2103–2109. https://doi.org/10.1007/s00216-019-02227-w.

This reference is a key reference because it allowed us to compare our results with those of other authors