Process Biochemistry 51 (2016) 1338–1347

Contents lists available at ScienceDirect

Process Biochemistry

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

Artificial intelligence approach based on near-infrared spectral data for monitoring of solid-state fermentation

Augusto César Barchi[a], Shuri Ito[a], Bruna Escaramboni[b], Pedro de Oliva Neto[b], Rondinelli Donizetti Herculano[c], Matheus Carlos Romeiro Miranda[c], Felipe José Passalia[d], José Celso Rocha[d], Eutimio Gustavo Fernández Núñez[a,*]

[a] Grupo de Engenharia de Bioprocessos, Departamento de Ciências Biológicas, Universidade Estadual Paulista 'Júlio de Mesquita Filho' Campus-Assis, Avenida Dom Antônio, 2100, 19806-900 Assis, SP, Brazil
[b] Laboratório de Biotecnologia Industrial, Departamento de Ciências Biológicas, Universidade Estadual Paulista 'Júlio de Mesquita Filho' Campus-Assis, Avenida Dom Antônio, 2100, 19806-900 Assis, SP, Brazil
[c] Instituo de Química—Araraquara, Universidade Estadual Paulista 'Júlio de Mesquita Filho' Campus-Araraquara, Rua Professor Francisco Degni, 55, 14800-900 Araraquara, SP, Brazil
[d] Laboratório de Matemática Aplicada, Departamento de Ciências Biológicas, Universidade Estadual Paulista 'Júlio de Mesquita Filho' Campus-Assis, Avenida Dom Antônio, 2100, 19806-900 Assis, SP, Brazil

ARTICLE INFO

Article history:
Received 25 March 2016
Received in revised form 23 June 2016
Accepted 15 July 2016
Available online 18 July 2016

Keywords:
Artificial neural network
Bioprocess monitoring
Chemometrics
Enzymes
NIR spectroscopy
Partial least squares

ABSTRACT

This work aimed to establish a chemometric technique for quantifying amylase and protease activities as well as protein concentration in aqueous extracts of *Rhizopus microsporus* var. *oligosporus* obtained via solid-state fermentation (SSF). The kinetics of four agro-industrial wastes (wheat bran, soybean meal, type II wheat flour and sugarcane bagasse) were studied for 144 h, along with two different sets of their ternary mixtures, at a constant fermentation time of 120 h, to obtain primary data (biochemical parameters as well as near-infrared (NIR) spectral data). Then, models such as artificial neural network (ANN) and partial least squares (PLS) were calibrated to predict biochemical parameters using the spectral data. Primary data and three methods of preprocessing data – first, second and third derivatives – were assessed as inputs for both chemometric tools. The third derivative, that is, spectral pre-processing plus an optimized ANN, showed the least relative errors (<8.3% ± 10.5%). The third-derivative spectrum was found to be suitable as the ANN input data for monitoring amylase and protease activities and protein concentration in the SSF under study. The proposed methodology can serve as a foundation for at-line sensor development and decrease the time and cost of bioprocess development using *Rhizopus microsporus* var. *oligosporus*.

1. Introduction

Solid-state fermentation (SSF) is the fermentation of solid substrates with low water content, but sufficient for the metabolism and growth of microorganisms. While the water content in classical submerged fermentation (SmF) is >95%, that in SSF varies between 40% and 80%. SSF often utilizes agricultural and food wastes as substrates [1,2]. SSF has been suggested as an alternative to SmF, as the former has several advantages such as a lower implementation cost and a decreased chance of contamination. Moreover, the associated products are generally more tolerant to pH and temperature variations, and their separation is not complex, as a rule [1,3,4]. Recently, the SSF has become more accepted by industrial and academic communities due to its wide range of applications, for instance, producing secondary metabolites and extracellular enzymes used in food, pharmaceutical and biofuel industries (i.e., amylases and proteases). Furthermore, with the growing demand for 'green processes' as alternatives of chemical processes in the manufacturing industry, SSF has garnered great interest [5].

However, the major limitations of solid-state bioprocesses include the control of operation parameters as well as measurements of growth and metabolic parameters, such as cell concentration, substrate consumption and production of biomolecules. These limitations are attributed to the heterogeneous nature of the substrate, which has a highly complex structure and nutrient composition [1,6]. Furthermore, unlike SmF, the SSF medium presents several moisture and temperature gradients, which can have a negative impact on the process efficiency [7].

* Corresponding author.
E-mail address: eutimiocu@yahoo.com (E.G. Fernández Núñez).

http://dx.doi.org/10.1016/j.procbio.2016.07.017
1359-5113/

On the other hand, spectroscopy techniques can be used to monitor several parameters in bioprocesses in three different ways: (1) off-line methods, which are based on analysis of samples transferred to the laboratory, away from the bioreactor; (2) at-line measurements, which also involve manual or automatic sampling, but the analyses are conducted in the vicinity of the bioreactor; and (3) online methods, which use sensors in various ways to provide real-time analyses. Spectroscopy techniques such as ultraviolet–visible (UV–vis), near-infrared (NIR), mid-infrared (MIR), Raman and fluorescence spectroscopy, with fibre optic cables, have been evaluated for their efficiency in bioprocess online monitoring (in situ) [8].

In particular, NIR spectroscopy is classified as a molecular spectroscopic method with a spectrum region at 700–2500 nm [9]. The vibrations associated with the X–H bond type, such as O-H, N-H, C-H and S-H, show a high dipole momentum, which is widely found in biological molecules and thus forms the underlying principle of NIR spectroscopy [10]. Currently, NIR spectroscopy is used in the pharmaceutical, agriculture, chemical and food industries for quality control from the raw material stage to the final products [11].

Chemometrics, a useful tool for bioprocess monitoring, utilizes mathematical and statistical methods to analyse data from chemical systems. This tool can be used to model and extract information from a large amount of data obtained monitoring with spectroscopy techniques [12]. For instance, in fermentative methods, with substantial overall data (input spectral data and related output biochemical parameters), techniques such as PLS and partial component regression (PCR) can be used to analyse a complex mixture to determine the quantities of minor compounds [13]. Artificial neural networks (ANNs) have emerged as an alternative to these multivariate techniques. This artificial intelligence approach is a modelling methodology based on the mechanisms of the human brain. This technology uses simple processing units, known as neurons, arranged in input, hidden and output layers. By learning functions, the ANN can improve the model results by refining the parameters based on the obtained errors [14,15]. ANNs can be used successfully for bioprocess monitoring as they are flexible in the use of either linear or non-linear functions (or a combination of both) [16,17].

The present work aimed to correlate NIR spectral data from the SSF samples with the total protein concentration as well as amylolytic and proteolytic activities using ANN and PLS models. This establishes a foundation for developing sensors for SSF monitoring. It also reduces the time and cost of kinetic studies of the biotransformation of agro-industrial wastes by *Rhizopus microsporus* var. *oligosporus*, as well as quantifying the yield over the course of the scale-up stages.

2. Materials and methods

2.1. Microorganism strain and substrates

The microorganism used in the present work, *R. microsporus* var. *oligosporus* (CCT 3762), was obtained from the Tropical Foundation of Research and Technology 'André Tosello', Campinas, SP, Brazil. The microorganism population was expanded in 3.9% (m/v) potato dextrose agar (PDA) medium using 250-mL Erlenmeyer flasks. The spore solution, for inoculation of SSF experiments, was obtained after 7 days of incubation (30 °C), and the biomass was washed with 0.01% (v/v) Tween 80 in an aqueous solution under gentle magnetic stirring. The SSFs were carried out using agro-industrial wastes as substrates. The four selected substrates were wheat bran (Moinho Nacional, Assis, SP, Brazil), soybean meal (Landtech Comércio e Indústria Ltda, Paraguaçu Paulista, SP, Brazil), type II wheat flour (Moinho Nacional, Assis, SP, Brazil) and sugarcane bagasse (Água Bonita, Tarumã, SP, Brazil). They were used individually or as a component of ternary mixtures.

Table 1
Experimental matrix corresponding to D-optimal mixture design for studying the influence of ternary mixtures comprising wheat bran (A), type II wheat flour (B) and sugarcane bagasse (C) on spectral data modification in extracts and their associated amylolytic and proteolytic activities as well as total protein concentration by means of *Rhizopus microsporus* var. *oligosporus* in solid-state fermentation.

Run	A [%]	B [%]	C [%]
1	75.00	25.00	0.00
2	60.00	33.33	6.67
3	30.00	50.00	20.00
4	80.00	0.00	20.00
5	50.00	50.00	0.00
6	70.00	16.67	13.33
7	47.50	37.50	15.00
8	82.50	12.50	5.00
9	55.00	25.00	20.00
10	30.00	50.00	20.00
11	50.00	50.00	0.00
12	100.00	0.00	0.00
13	80.00	0.00	20.00
14	90.00	0.00	10.00

Table 2
Experimental matrix corresponding to D-optimal mixture design for studying the influence of ternary mixtures comprising type II wheat flour (B), sugarcane bagasse (C) and soybean meal (D) on spectral data modification in extracts and their associated amylolytic and proteolytic activities as well as total protein concentration by means of *Rhizopus microsporus* var. *oligosporus* in solid-state fermentation.

Run	B [%]	C [%]	D [%]
1	48.33	13.33	38.33
2	50.00	15.00	35.00
3	40.00	20.00	40.00
4	48.33	18.33	33.33
5	50.00	15.00	35.00
6	50.00	10.00	40.00
7	45.83	18.33	35.83
8	50.00	20.00	30.00
9	43.33	18.33	38.33
10	45.00	20.00	35.00
11	50.00	10.00	40.00
12	40.00	20.00	40.00
13	46.67	16.67	36.67
14	45.00	15.00	40.00
15	50.00	20.00	30.00

2.2. Fermentation process

The SSFs were performed in 250-mL Erlenmeyer flasks in two different ways: (1) the substrate was composed of 10 g of a single agro-industrial waste, or (2) the substrate contained the same mass of a ternary mixture with the composition in line with runs included in two D-optimal mixture designs (Design-Expert 6.0.1 version, Stat-Ease, Inc., Minneapolis, MN, USA). The ternary mixtures are listed in Tables 1 and 2. In the first case, the fermentations were conducted as kinetic studies. They were conducted in triplicate and monitored at 24-h intervals for 144 h. In the second case, the duration was 120 h. These two different approaches generated a wide range of data. In both cases, the substrate was humidified using an aqueous micronutrient solution composed of 1.25% (m/m) ammonium sulphate, 0.25% (m/m) potassium phosphate and 1.25% (m/m) urea, so as to obtain 55% (m/m) moisture [18]. After substrate sterilization (121 °C, 1 atm, 30 min), the inoculation was aseptically performed with 1×10^6 spores (concentration of spore suspension: 2.27×10^7 spores/mL) of *R. oligosporus* per gram of the fresh substrate. Subsequently, the Erlenmeyer flasks were incubated at 30 °C (Tecnal TE-421, Piracicaba, Brazil).

2.3. Enzymatic extraction

The enzyme extraction was performed by adding 100 mL of distilled water (10 mL H_2O per gram of the fresh substrate) in the fermentation system. Then, the flasks were agitated at 18.85 rad/s for 30 min at 28 °C in an orbital shaker (Tecnal TE-421, Piracicaba, Brazil). Sequentially, the contents were filtered and then biochemically analysed.

2.4. Proteolytic activity

The proteolytic activity in aqueous extracts was measured according to Leighton's method [19] with some modifications. First, a 500-μL aliquot of the enzymatic extract was diluted threefold in distilled water. Diluted extract samples (150 μL) were added to 250 μL of azocasein solution (10 mg/mL in 0.05 M acetate buffer, pH = 5.5). Consecutively, the samples were placed in a water bath at 50 °C for 30 min. Then, 1000 μL of 10% trichloroacetic acid (TCA) (m/v) was added to halt the enzymatic reaction, and the contents of the test tubes were centrifuged to precipitate proteins. Next, 1000 μL of the resulting supernatant was mixed with 1000 μL of 1 M sodium hydroxide solution. The absorbance of the final mixture was measured in duplicate at 440 nm using a spectrophotometer (model UV-M51, Bel Engineering, Piracicaba, Brazil).

The blank was performed replacing the sample by distilled water. Some components of the diluted extract may interfere in the absorbance measurement, overvaluing the proteolytic activities in the extracts. To prevent any interference, an enzymatic control was performed with an additional test using 150 μL of denatured enzymatic extract (5 min in boiling bath water).

One unit (U) of proteolytic enzyme activity was defined as the amount of enzyme needed to change the absorbance by 0.01 per reaction minute. The calculation of enzymatic activity was performed according to Eq. (1) and expressed as enzymatic activity unit per gram of the fresh substrate (U/g):

$$PA = \frac{(Abs - Abs_{dn}) \cdot V_e \cdot D_F}{V_s \cdot t \cdot m_{sfw} . F_c} \quad (1)$$

where PA is the proteolytic activity (U/g), Abs is the absorbance corresponding to the sample of the diluted enzymatic extract at 440 nm, Abs_{dn} is the absorbance corresponding to the sample of denatured extract at 440 nm, V_e is the volume of the total extract (100 mL), D_F is the dilution factor (3), V_s is the diluted sample volume (150 μL), t is the reaction time (30 min), m_{sfw} is the solid substrate mass corresponding to the total extract volume (10 g) and F_c is the conversion factor absorbance in enzymatic units (0.01(1/[U × min])).

2.5. Amylolytic activity

The amylolytic activity was measured by the quantification of released sugar with reducing groups, using the 3,5-dinitrosalicylic acid (DNS) method [20]. Test tubes containing 650 μL of 0.5% (m/v) starch solution in 0.05 M acetate buffer (pH = 5.5) were placed in a water bath with a temperature control (Tecnal TE-056, Piracicaba, Brasil) to reach 60 °C. To each of these test tubes, 100 μL of the enzymatic extract diluted 20-fold in water was added. After 10 min at 60 °C, 500 μL of DNS reagent was added. The tubes with the reaction mixtures were sealed and placed in a boiling water bath for 5 min, and then cooled in an ice bath. Distilled water (3750 μL) was added to test tubes, and the absorbance was measured in duplicate at 540 nm using a UV–vis spectrophotometer (model UV-M51, Bel Engineering, Piracicaba, Brazil).

As a reference sample (blank), the total reaction mixture volume (750 μL) was replaced by the same volume of distilled water. As an enzyme control was also needed, 100 μL of the enzymatic extract was added to 650 μL of distilled water. The same procedure described earlier was followed. One unit of amylase activity (U) was defined as the amount of enzyme needed to release 1 μmol of glucose per minute of the reaction.

Amylolytic activity (*AA*) was expressed in enzymatic activity unit per gram of the fresh substrate (U/g) and was calculated by Eq. (2):

$$AA = \frac{C_{rg} \cdot V_e \cdot D_F}{t_r \cdot V_s \cdot m_{sfw}} \quad (2)$$

where C_{rg} is the release glucose concentration, V_e is the extract volume (100 mL), D_F is the dilution factor (20), t_r is the reaction time, V_s is the diluted extract sample volume (100 μL) and m_{sfw} is the mass of the fresh substrate (10 g).

2.6. Total protein concentration

The total protein concentration was determined by the Bradford method [21]. A 1000-μL aliquot of the working solution, which contains Coomassie Brilliant Blue G-250 (Bradford reagent), was added to 100 μL of the enzymatic extract. The reaction mixture was protected from natural light for 5 min, and then its absorbance was measured in duplicate at 595 nm using a spectrophotometer (model UV-M51, Bel Engineering, Piracicaba, Brazil).

The reference sample (blank) was prepared by replacing the enzymatic extract with the same volume of distilled water. The total protein concentration was calculated using Eq. (3) and measured in micrograms per millilitre (μg/mL). Previously, a calibration curve was elaborated using aqueous solutions of bovine serum albumin with different known concentrations (range 0–250 μg/mL):

$$C_p = \frac{Abs - b}{a} \cdot D_F \quad (3)$$

where C_p is the total protein concentration of the extract (μg/mL), Abs is the measured absorbance, a is the calibration curve slope (mL/μg), b is the calibration curve intercept and D_F is the dilution factor.

2.7. NIR spectral data acquisition

The enzymatic extracts (samples) were maintained at −20 °C (within a time frame when spectral changes are not observed) until the spectrophotometric analysis. For each extract, 3.5 mL of the sample was analysed in a UV–NIR spectrophotometer (Perkin Elmer, Lambda 950, EUA) using a quartz cuvette and air as reference. The spectral analysis was performed at a wavelength range of 700–1800 nm to obtain transmittance data. In general, the samples were not diluted, but sample dilutions were necessary in some cases (dilution factor of 2.3). After the transmittance spectrum was obtained, the data were saved in a Microsoft Excel 2007 file (Microsoft Corporation, Redmond, WA, USA) for further analyses.

2.8. Preprocessing of data

Primary data from the NIR spectra were preprocessed using the SIMCA 14 software demonstration version (Umetrics, Umeå, Sweden). Three preprocessing techniques were assessed – first, second and third derivatives – because overlapping problems had to be minimized [22]. Other techniques were also considered (data not shown), but prediction results were not suitable. The data matrices derived from preprocessing, as well as the original data, were used for calibration of the ANN and PLS models.

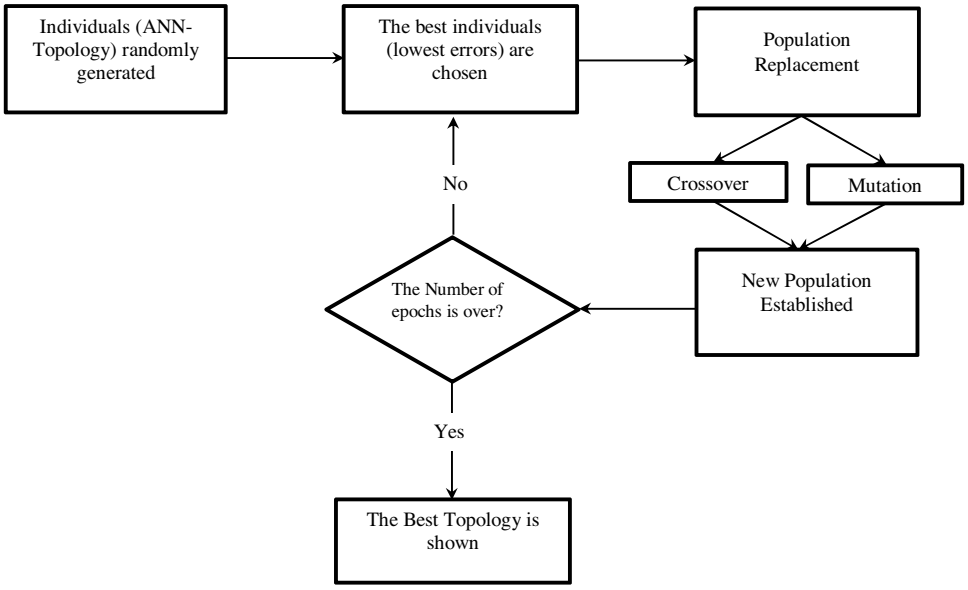

Fig. 1. Flow chart representing algorithm to ANN architecture optimization based on genetic algorithm methodology.

2.9. ANN modelling

The authors developed a code for topology optimization of the ANN based on a genetic algorithm (GA) in the MATLAB platform (MATLAB R2012a, Mathworks Inc., Natick, MA, USA). The present code was programmed using a similar algorithm developed by our group [23]. The steps of the GA-ANN algorithm are described in a flow chart (Fig. 1). The main input parameters for this algorithm were as follows: population size (200 individuals); number of neurons in hidden layer 1 (4–100), 2 (4–100) and 3 (1–10); transfer functions for hidden layers and for output layer (saturating linear transfer function [satlin], log-sigmoid transfer function [logsig], hyperbolic tangent sigmoid transfer function [tansig] and linear transfer function [purelin]); training function (Resilient backpropagation [trainrp]); mutation rate (5%); survival rate (20%); and number of epochs (20).

The 101 samples, which correspond to the overall data obtained, were randomly divided into training (70%), validation (15%) and testing (15%) sets. This partitioning of the original database was done according to criterion previously reported for equivalent problems (spectral data used as input data for ANN calibration) [24,25]. At the end of each epoch, the ANN topology with the least errors (test error) was chosen. The input 'neurons parameter' was fixed as 1102 when the original data were used (associated with transmittance in the 700–1800-nm range) and 1088 for preprocessing data (associated with transmittance in the 707–1793-nm range) plus another neuron associated with the dilution factor. One neuron was present in the output layer, corresponding to values of individual response variable under consideration (amylase and protease activities and total protein concentration in extracts). In summary, four ANNs were assessed for each of the three biochemical parameters. The four evaluated ANNs corresponded to different ANN input data, primary spectral data, and three other sets of spectral data derived from preprocessing techniques (first, second and third derivatives).

2.10. PLS modelling

The matrix X, comprising independent variables, included transmittance data within the spectrum range under study (700–1800 nm) in the original or preprocessed form with better prediction results obtained by ANN modelling. The dilution factors were also considered. Response matrix Y was defined by a column vector comprising values of amylolytic and proteolytic activities or total protein concentration, depending on the model to be adjusted. PLS modelling was performed in SIMCA 14 demonstration version (Umetrics, Umeå, Sweden). All samples were utilized to adjust the PLS models. The quality of prediction for these models was assessed by coefficients of correlation (R) and prediction (Q^2).

2.11. Statistical analysis

The selection of the best set of spectral data (original or any preprocessed alternative) to ensure the least prediction error associated with each response variable under consideration in both correlation techniques was defined by one-way analysis of variance and subsequently Tukey's test for multiple comparison of the set of means. This procedure was performed to detect differences among means of errors. Statistical decisions were made with the level of significance at 95% ($\alpha = 0.05$).

The comparison between PLS and ANN (best data preprocessing and ANN architecture) models was performed considering the absolute error. For this purpose, a one-tail Student's *t*-test was conducted with $\alpha = 0.05$.

All statistical tests were performed in the software BioEstat 5.0 (Instituto de Desenvolvimento Sustentável, Mamirauá, Brazil).

3. Results

3.1. Kinetic studies: enzyme activities and protein concentration

The amylase produced from SSF depended on the agro-industrial wastes used. Among the assessed wastes, wheat bran showed better results (Fig. 2A). By contrast, sugarcane bagasse showed the lowest values for this group of enzymes. The fermentation times for maximum amylolytic activity values were not the same among the residues under study. The maximum amylase production for wheat bran was confirmed at 120 h. Then, a decrease in amylolytic enzyme activities was observed. A similar expression profile was confirmed for sugarcane bagasse as well, which was similar to those reported for different microorganisms [26]. Moreover, the kinetics profiles of amylase production associated with type II wheat flour and soybean meal differed from others; between them, the highest amylase activities were detected at a

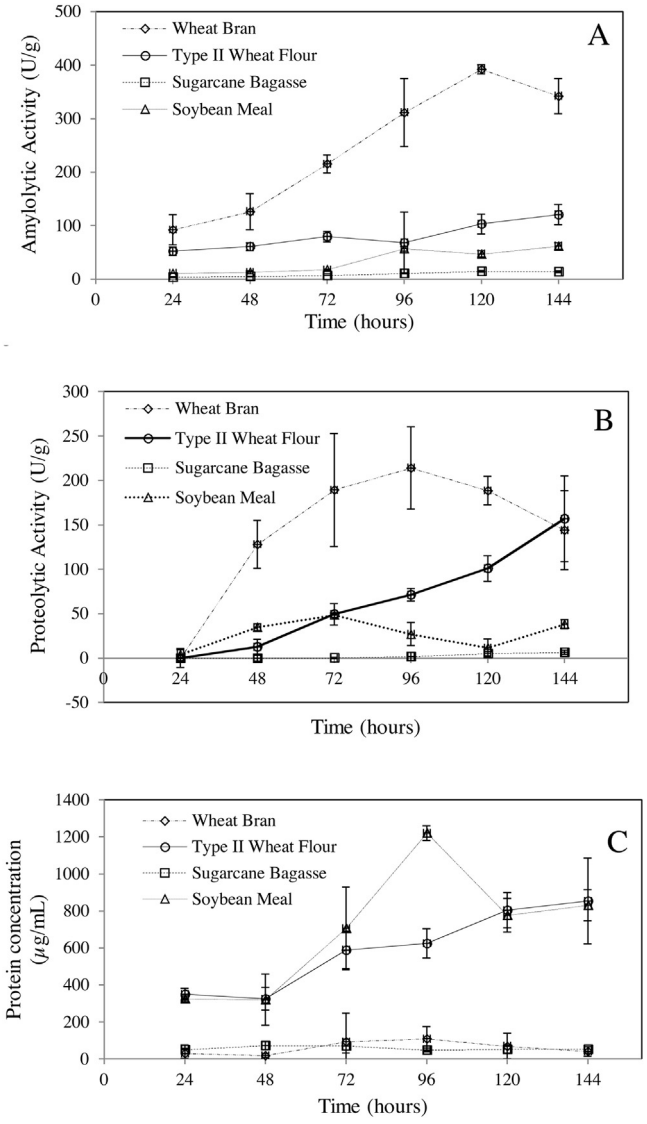

Fig. 2. Kinetic studies for enzyme and protein production by *Rhizopus microsporus* var. *oligosporus* (CCT 3762) using different agro-industrial wastes as substrates. A: Amylolytic enzyme. B: Proteolytic enzyme. C: protein concentration. Symbols and bars represent the averages and standard deviations for three determinations of each monitored parameter.

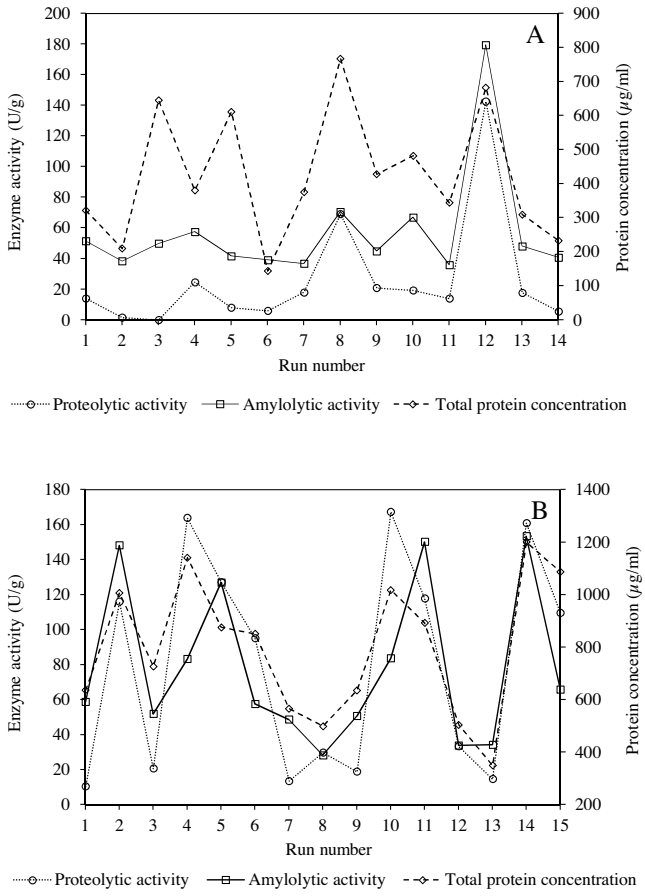

Fig. 3. Response variables, amylolytic and proteolytic activities and protein concentration in aqueous extract from experimental runs associated with mixture designs previously defined. A: Results corresponding to D-optimal design described in Table 1 (ternary mixtures comprising wheat bran, type II wheat flour and sugarcane bagasse). B: Results corresponding to D-optimal design described in Table 2 (ternary mixtures comprising type II wheat flour, sugarcane bagasse and soybean meal).

fermentation time of 144 and 96 h, respectively. The high levels of amylase activity for wheat bran compared with other assessed agro-industrial wastes can be attributed to the physical and chemical (protein content of 15% and high levels of polysaccharides (56%), most of them being starch) properties of this substrate [27,28].

As noted in the study of amylase production, the patterns of proteolytic enzyme production were not similar for the wastes under study. Maximum protease production was again confirmed for wheat bran and minimum for sugarcane bagasse (Fig. 2B). However, the differences in protease expression among agro-industrial wastes were not as significant as those observed in the study of amylases (Fig. 2A). For example, the protease activity for wheat bran was 1.49-fold higher than the corresponding activity for type II wheat flour, whereas this ratio was 3.25 for amylase activity (Fig. 2A and B). The fermentation times associated with maximum protease activity as a rule were similar to or shorter than those values observed for amylase activity.

The protein profiles in aqueous extracts from fermented agro-industrial wastes were significantly different from those reported for amylase and protease activities. The highest values of protein concentration in the extract were related to soybean meal, followed by type II wheat flour, wheat bran and sugarcane bagasse (Fig. 2C). In the case of protein profiles, soybean meal and type II wheat flour showed similar patterns corresponding to the amylase and protease activities. The protein concentration ratio between soybean meal and wheat bran was 11.3-fold at the concentration peaks (Fig. 2C). The highest values of protein concentration obtained for soybean meal, despite its poor amylolytic and proteolytic enzyme expression, can be attributed to the high protein content (49.24%) in this substrate [27].

3.2. Mixture experimental designs

The values corresponding to response variables, which characterize the aqueous extracts for both ternary agro-industrial waste mixtures included in the present work, are presented in Fig. 3. No statistical modelling was performed to determine the influence of residue proportion on amylase and protease activities, because it was outside the scope of this paper. Similar results were obtained for protein concentration (Tables 1 and 2, Figs. 2 and 3). However, a significant difference was noted between enzyme activities associated with similar fermentation experiments in previous kinetic studies. This finding was more remarkable for amylolytic enzyme activity and could be explained by changes in the sporulation levels (Fig. 4). In a previous study, five levels of sporulation for *R.*

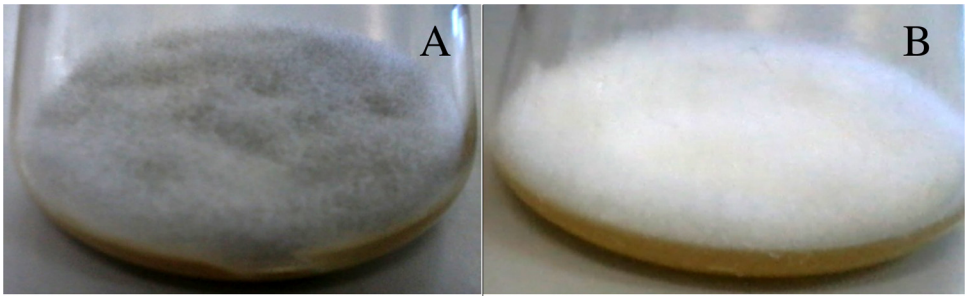

Fig. 4. Different sporulation levels in *Rhizopus microsporus* var. *oligosporus* (CCT 3762) at the same growth conditions in an experiment included in kinetic studies (A) and equivalent run of mixture design (B).

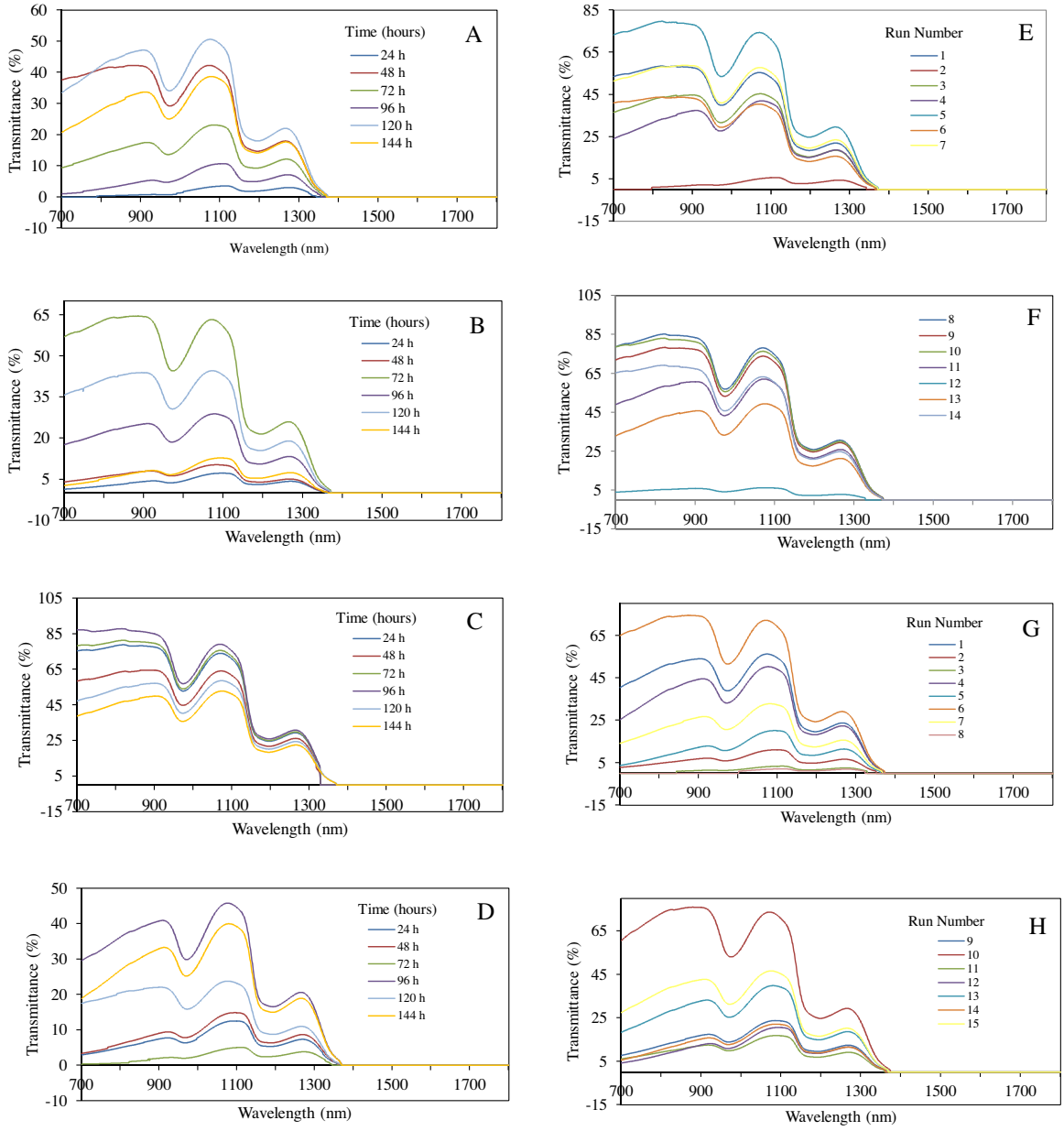

Fig. 5. NIR spectral data derived from aqueous extracts of individual or ternary mixtures of agro-industrial wastes. A: Samples corresponding to kinetic study of wheat bran (each spectrum represents the average for three repetitions). B: Samples corresponding to kinetic study of type II wheat flour (each spectrum represents the average for three repetitions). C: Samples corresponding to kinetic study of sugarcane bagasse (each spectrum represents the average for three repetitions). D: Samples corresponding to kinetic study of soybean meal (each spectrum represents the average for three repetitions). E: Samples corresponding to experimental runs (1–7) from D-optimal design described in Table 1. F: Samples corresponding to experimental runs (8–15) from D-optimal design described in Table 1. G: Samples corresponding to experimental runs (1–7) from D-optimal design described in Table 2. H: Samples corresponding to experimental runs (8–15) from D-optimal design described in Table 2.

microsporus were reported. In most strains of *R. microsporus* var. *oligosporus*, amylase production is associated with high sporulation levels [29].

3.3. Spectral data

The spectral data corresponding to the aqueous extracts from kinetic studies are presented in Fig. 5A–D. The spectral profiles are similar in the overall shape, but they differ with the type of agro-industrial waste and fermentation time. The spectral profiles are characterized by two valleys (around 970 and 1200 nm) and three peaks (around 880, 1080 and 1270 nm) of transmittance. The spectral data corresponding to aqueous extracts from ternary mixture studies are presented in Fig. 5E–H. The spectral patterns corresponded to the proportions of agro-industrial residues.

The preprocessed spectra from the first, second and third derivative were compared with the original spectrum. The spectrum pattern was found to increase in complexity with higher derivatives (Fig. 6).

3.4. GA-ANN modelling

The values of the correlation coefficients (R) for training, validation, test and overall stages as well as averages for absolute errors (absolute difference between observed and predicted values) for each ANN with optimized topology are reported in Table 3.

The influence of the preprocessing technique for primary spectral data on the decrease of absolute errors was only statistically significant for the protein concentration, with the first and third derivatives improving the prediction quality in relation to the original data (Tukey's test, Table 3). However, preprocessing of the spectral data with the third derivative showed the lowest average and highest overall correlation coefficients for the three responses under consideration. The optimal ANN topologies for these input data sets are presented in Table 4. As a rule, two hidden layers and the training function 'trainrp' were common for all three responses, but the distribution of neurons in layers and the transfer functions were dissimilar. The need for two hidden layers for three extract parameters is common in complex problems [30].

The amylolytic activities ranged from 3.35 to 399.86 U/g. The average of absolute error (12.03 ± 17.82 U/g) was relatively low, taking into account the size of the range (difference between maximum and minimum value of the range, 396.51 U/g). The relative error considering the range for this response variable was $3.0\% \pm 4.5\%$. In the case of protease activity, the relative error was $8.3\% \pm 10.5\%$, whereas this prediction quality index was $5.1\% \pm 7.4\%$ for protein concentration.

3.5. PLS modelling

The preprocessing of spectral data, by the third derivative, significantly increased the correlation coefficient associated to linear regression between observed and predicted values of response variables (values nearest to 1, better fit to the experimental data) with respect to primary data in the case of PLS modelling (Table 5). However, the prediction coefficient was not suitable ($Q^2 < 0.5$) for all extract parameters, even with data preprocessing. According to one-way analysis of variance and subsequently Tukey's test, the absolute error was lower for proteolytic activity and protein concentration associated with preprocessed data ($p < 0.05$). No improvement was observed for amylase activity through data preprocessing ($p = 0.0639$).

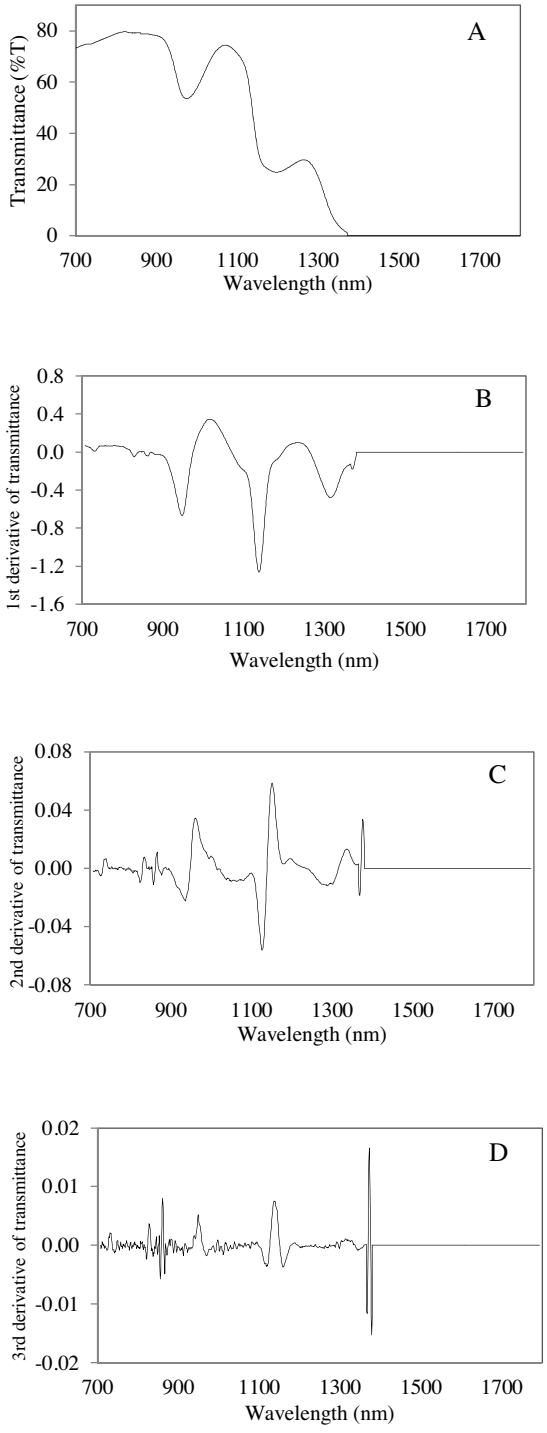

Fig. 6. Changes in the spectral profile caused by preprocessing data for a generic sample included in input data set. Where A is the original transmittance spectra; B is the first-derivative transmittance spectra; C is the second transmittance and D is the third transmittance spectra.

3.6. Comparison between modelling approaches

On comparing the absolute errors associated with ANN (Table 3) and PLS (Table 5) models, using the third derivative as spectral data pre-processing, ANN errors were found to be lesser than PLS errors for amylase activity and protein concentration ($p < 0.0001$ for one-tail Student's t-test). However, no difference was noted between the two modelling approaches for protease activity ($p = 0.069$).

Table 3
Artificial neural network models with best network topology defined by computational algorithm for each response variable being considered and their corresponding correlation coefficient for linear regression between predicted and observed values as well as their absolute error.

Response variable	Data preprocessing	*R* (training stage)	*R* (validation stage)	*R* (test stage)	*R* (overall)	Absolute error (average ± standard deviation)
Amylolytic Activity (U/g)	Primary data	0.9833	0.7971	0.7472	0.9302	19.31 ± 30.41 U/g [(a)*]
	1 st derivative	0.8853	0.7908	0.7506	0.8361	34.53 ± 41.42 U/g[(b)]
	2nd derivative	0.6408	0.8468	0.7505	0.6723	49.75 ± 54.36 U/g[(c)]
	3rd derivative	0.9954	0.9358	0.8939	0.9755	12.03 ± 17.82 U/g[(a)]
Proteolytic Activity (U/g)	Primary data	0.8981	0.7355	0.8221	0.8446	23.49 ± 28.96 U/g[(a)]
	1 st derivative	0.7687	0.7717	0.7423	0.7648	29.30 ± 33.00 U/g[(a)]
	2nd derivative	0.7572	0.7887	0.7874	0.7657	28.77 ± 33.39 U/g[(a)]
	3rd derivative	0.8755	0.918	0.8838	0.8677	21.19 ± 26.83 U/g[(a)]
Protein concentration (μg/mL)	Primary data	0.6556	0.656	0.8124	0.6632	382.81 ± 386.16 μg/mL[(b)]
	1 st derivative	0.8021	0.8661	0.7168	0.7936	153.06 ± 159.57 μg/mL[(a)]
	2nd derivative	0.6558	0.5744	0.6757	0.6416	392.7 ± 334.83 μg/mL[(b)]
	3rd derivative	0.9895	0.8369	0.8721	0.9529	64.04 ± 92.68 μg/mL[(a)]

* Means with different letters were significantly different ($p < 0.05$, ANOVA, Tukey's test).

Table 4
Parameters of optimal artificial neural network topologies for best preprocessing data associated with each response variable under consideration.

Parameters	Amylolytic Activity (U/g)	Proteolytic Activity (U/g)	Protein concentration (μg/mL)
Hidden layers	2	2	2
Distribution of neurons in layers	1088-82-24-1	1088-16-8-1	1088-84-29-1
Transfer function	logsig-logsig-tansig	logsig-tansig-tansig	logsig-logsig-tansig
Training function	trainrp	trainrp	trainrp
Spectral data preprocessing	3rd derivative	3rd derivative	3rd derivative

Table 5
PLS modelling for each response variable under study and their respective values of correlation coefficient (*R*, linear regression between predicted and observed values), prediction coefficient (Q^2) and absolute error.

Response variable	Spectral data preprocessing	Number of significant principal components	Correlation coefficient (*R*)	Prediction coefficient (Q^2)	Absolute error (Average ± Standard deviation)
Amylolytic Activity (U/g)	Primary data	5	0.7688	0.371	45.89 ± 41.73 U/g
Proteolytic Activity (U/g)		5	0.712	0.274	35.59 ± 32.055 U/g
Protein concentration (μg/mL)		2	0.4099	0.118	262.47 ± 196.99 μg/mL
Amylolytic Activity (U/g)	3rd derivative	3	0.8643	0.3842	36.21 ± 32.71 U/g
Proteolytic Activity (U/g)		3	0.8701	0.3317	26.42 ± 20.77 U/g
Protein concentration (μg/mL)		3	0.8927	0.483	128.62 ± 98.39 μg/mL

4. Discussion

The challenge in the present work was to establish a correlation between NIR spectral data and amylolytic and proteolytic activities as well protein concentration. The aim was to design a chemometrics ANN-based technique that could replace conventional analyses, which have many steps, require expensive reagents and are time consuming [31].

In general, the NIR spectra for all samples showed modification of one basic profile, which is associated with the NIR spectrum for water molecules [32]. The observed valleys in the NIR transmittance spectra (around 970 and 1200 nm) match with previously reported peaks in the NIR absorbance spectra for water. The 970-nm peak is related to a combination band (2 × symmetric OH stretching mode (ν_1) + antisymmetric OH stretching mode (ν_3)) and the 1200-nm peak to another combination band (ν_1 + OH bending mode (ν_2) + ν_3) [33,34]. Similar spectral data were obtained for other systems; thus, small changes in the spectra were highlighted to quantify chemical compounds in aqueous solutions [35]. For this reason, all spectral data of the samples included in this work were considered to predict the most appropriate system. Nevertheless, data pre-processing techniques were required to optimize the predictive and generalization power of chemometric models. This type of technique is widely used to increase the applicability of chemosensors based on NIR spectra [14]. The most classical spectral data pretreatments include normalizations, derivatives and smoothing [36]. Specifically, derivatization of spectral data, as a rule, has the advantages of sharpening peaks and resolving overlapping bands to some extent, although this kind of pretreatment amplifies the noise of the data [37]. However, the best data preprocessing approach for NIR spectral data must be selected empirically, using a trial-and-error methodology, which is a major drawback [38].

In this study, the application of third-derivative preprocessing was found to cause marked changes in the NIR spectrum, which allowed the GA-ANN model to detect subtle differences among the samples. Moreover, in the literature, ANN is considered suitable for biocompound estimations when the overall correlation coefficient is >0.89 [39]. Spectral data pre-processing with the third derivative met this criterion for amylase activity and protein concentration; the corresponding value for protease activity was almost similar to but lower than the reference value for the correlation coefficient.

The superior or similar prediction performance of the ANN modelling approach to PLS models (even without performing validation-test of PLS models and including all samples in calibration stage of PLS models) can be explained by the limited linearity of biological systems. Therefore, ANNs are best suited for experimental data compared to multivariate techniques such as PLS [25,39]. PLS was assessed in the present work because it is relatively

simple and widely used in chemometrics associated with NIR spectroscopy. Thus, PLS results were used as a reference for comparison with the GA-ANN model proposed.

As mentioned in Section 3.4, the GA-ANN model successfully predicted the response variables. Furthermore, the set of samples, which was used to train, validate and test the model derived from SSF systems, differed in the substrate composition, fermentation time and sporulation levels. This helped ensure the comprehensiveness of the proposed ANN model. However, this model is only useful within the experimental domain considered. For this reason, this calibrated ANN did not provide good results with different substrates and microorganisms. We recommend calibrating new ANN models with particular conditions or establishing a collaborative database to widen their use in both industrial and academic research.

This work establishes the foundation for developing at-line sensors to determine simultaneously two types of enzymes in SSF with *R. microsporus* var. *oligosporus*. This chemosensor would be extremely useful for pilot-scale equipment to understand better the bioprocess and to assess different scale-up criteria, as well as for designing successful commercial-scale bioreactors. Studies on at-line sensors for simultaneous enzyme determination in SSF are scarce.

5. Conclusion

In conclusion, it is worth noting that ANN in combination with the third derivative as NIR spectral data preprocessing is suitable for monitoring amylase and protease activities and protein concentration in SSF of *R. microsporus* var. *oligosporus* in a wide range of experimental conditions, including four individual agro-industrial wastes, a ternary combination of these wastes, various fermentation times and different sporulation levels of the microorganism under study. For this reason, this chemometric technique can serve as the foundation for developing at-line chemosensors and can decrease time and cost of the bioprocess development, and also make these activities environmentally friendly (exempt for chemical reagents). To confirm the results of this work and before developing at-line chemosensors, validation experiments at a larger scale are needed.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank Fundação para o Desenvolvimento da Unesp (Fundunesp) for the scientific grant (0312/001/14-Prope/CDC) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for Scientific Initiation Scholarship (14/06447-0). The first author is grateful for the dedication of his former professor, Orlando da Mota Freitas. The corresponding author acknowledges his wife, Relma, and daughters, Giovanna and Paola, for the inspiration to write this manuscript.

References

[1] M. Shuler, F. Kargi, Bioprocess Engineering: Basic Concepts, 2nd ed., Prentice Hall PTR, Upper Saddle River, United States of America, 2002, pp. 276–278.
[2] L. Thomas, C. Larroche, A. Pandey, Current developments in solid-state fermentation, Biochem. Eng. J. 81 (2013) 146–161.
[3] U. Böhmer, S. Frömmel, T. Bley, M. Müller, K. Frankenfeld, P. Miethe, Solid-state fermentation of lignocellulotic materials for the production of enzymes by the white-rot fungus *Trametes hirsuta* in a modular bioreactor, Eng. Life Sci. 11 (4) (2011) 395–401.
[4] I. Haq, H. Mukhtar, Biosynthesis of proteases by *Rhizopus oligosporus* IHS_{13} in low-cost medium by solid-state fermentation, J. Basic Microbiol. 44 (4) (2004) 280–287.
[5] R. Singhania, Production of celluloytic enzymes for the hydrolysis of lignocellulosic biomass, in: A. Pandey, C. Larroche, S.C. Ricke, C. Dussap, E. Gnansounou (Eds.), Biofuels: Alternative Feedstocks and Conversion Processes, Elsevier, San Diego, United States of America, 2011, pp. 177–201.
[6] A. Pandey, Solid-state fermentation, Biochem. Eng. J. 13 (2–3) (2003) 81–84.
[7] C.S. Farinas, G.L. Vitcosque, R.F. Fonseca, V.B. Neto, S. Couri, Modeling the effects of solid state fermentation operating conditions on endoglucanase production using an instrumented bioreactor, Ind. Crops Prod. 34 (1) (2011) 1186–1192.
[8] T. Becker, B. Hitzmann, K. Muffler, R. Pörtner, K.F. Reardon, F. Stahl, R. Ulber, Future aspects of bioprocess monitoring, in: R. Ulber, D. Sell (Eds.), White Biotechnology, Springer, Berlin, Germany, 2007, pp. 249–293.
[9] G. Macaloney, J.W. Hall, M.J. Rollins, I. Draper, K.B. Anderson, J. Preston, B.G. Thompson, B. McNeil, The utility and performance of near-infra red spectroscopy in simultaneous monitoring of multiple components in a high cell density recombinant *Escherichia coli* production process, Bioprocess. Eng. 17 (3) (1997) 157–167.
[10] M. Scarff, S.A. Arnold, L.M. Harvey, B. McNeil, Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends, Crit. Rev. Biotechnol. 26 (1) (2006) 17–39.
[11] R.A. Forbes, M.L. Persinger, D.R. Smith, Development and validation of analytical methodology for near-infrared conformance testing of pharmaceutical intermediates, J. Pharm. Biomed. Anal. 15 (3) (1996) 315–327.
[12] N.D. Lourenço, J. a Lopes, C.F. Almeida, M.C. Sarraguça, H.M. Pinheiro, Bioreactor monitoring with spectroscopy and chemometrics: a review, Anal. Bioanal. Chem. 404 (4) (2012) 1211–1237.
[13] S. Vaidyanathan, B. McNeil, Near Infrared spectroscopy—a panacea in pharmaceutical bioprocessing? Eur. Pharm. Rev. 3 (1998) 43–48.
[14] R. Luttmann, D.G. Bracewell, G. Cornelissen, K.V. Gernaey, J. Glassey, V.C. Hass, C. Kaiser, C. Preusse, G. Striedner, C.F. Mandenius, Soft sensors in bioprocessing: a status report and recommendations, Biotechnol. J. 7 (8) (2012) 1040–1048.
[15] Q. Ding, G.W. Small, M.A. Arnold, Evaluation of nonlinear model building strategies for the determination of glucose in biological matrices by near-infrared spectroscopy, Anal. Chim. Acta 384 (3) (1999) 333–343.
[16] J. Glassey, G.A. Montague, A.C. Ward, B.V. Kara, Enhanced supervision of recombinant E. coli fermentation via artificial neural networks, Process Biochem. 29 (5) (1994) 387–398.
[17] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (1–2) (2010) 14–32.
[18] B. Escaramboni, P. Oliva-Neto, Brazilian Patent (2014), N° BR 10-2014-031591-8.
[19] T.J. Leighton, R.H. Doi, R.A.J. Warren, R.A. Kelln, The relationship of serine protease activity to RN polymerase modification and sporulation in bacillus subtilis, J. Mol. Biol. 76 (1) (1973) 103–122.
[20] G.L. Miller, Use of dinitrosalicylic acid reagent for determination of reducing sugar, Anal. Chem. 31 (3) (1959) 426–428.
[21] M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1–2) (1976) 248–254.
[22] D.M. Souza, F.F. Guimarães, Application of multivariate calibration and artificial intelligence in the analysis of infrared spectra to quantify organic matter in soil samples, Quim. Nova 35 (9) (2012) 1738–1745.
[23] M.B. Takahashi, J.C. Rocha, E.G. Fernández-Núñez, Optimization of artificial neural network by genetic algorithm for describing viral production from uniform design data, Process Biochem. 51 (3) (2016) 422–430.
[24] C. Cevoli, A. Gori, M. Nocetti, L. Cuibus, M.F. Caboni, A. Fabbri, FT-NIR and FT-MIR spectroscopy to discriminate competitors, non compliance and compliance grated Parmigiano Reggiano cheese, Food Res. Int. 52 (1) (2013) 214–220.
[25] M.B. Takahashi, J. Leme, C.P. Caricati, A. Tonso, E.G. Fernández Núñez, J.C. Rocha, Artificial neural network associated to UV/vis spectroscopy for monitoring bioreactions in biopharmaceutical processes, Bioprocess Biosyst. Eng. 38 (6) (2015) 1045–1054.
[26] A. Kunamneni, K. Permaul, S. Singh, Amylase production in solid state fermentation by the thermophilic fungus *Thermomyces lanuginosus*, J. Biosci. Bioeng. 100 (2) (2005) 168–171.
[27] R.J.S. de Castro, T.G. Nishide, H.H. Sato, Production and biochemical properties of proteases secreted by *Aspergillus niger* under solid state fermentation in response to different agroindustrial substrates, Biocatal. Agric. Biotechnol. 3 (4) (2014) 236–245.
[28] M. Prückler, S. Siebenhandl-Ehn, S. Apprich, S. Höltinger, C. Haas, E. Schmid, W. Kneifel, Wheat bran-based biorefinery 1: Composition of wheat bran and strategies of functionalization, LWT—Food Sci. Technol. 56 (2) (2014) 211–221.
[29] S. Dolatabadi, G. Walther, A.H.G. Gerrits Van Den Ende, G.S. de Hoog, Diversity and delimitation of *Rhizopus microsporus*, Fungal Divers. 64 (1) (2014) 145–163.
[30] M. Pirdashti, S. Curteanu, M.H. Kamangar, M.H. Hassim, M.A. Khatami, Artificial neural networks: applications in chemical engineering, Rev. Chem. Eng. 29 (4) (2013) 205–239.

[31] Y. Horikawa, M. Imai, K. Kanai, T. Imai, T. Watanabe, K. Takabe, Y. Kobayashi, J. Sugiyama, Line monitoring by near-infrared chemometric technique for potential ethanol production from hydrothermally treated *Eucalyptus globulus*, Biochem. Eng. J. 97 (2015) 65–72.
[32] A. Inoue, K. Kojima, Y. Taniguchi, K. Suzuki, Near-infrared spectra of water and aqueous electrolyte solutions at high pressures, J. Solution Chem. 13 (11) (1984) 811–823.
[33] S. Garrigues, M. de la Guardia, Methods for the vibrational spectroscopy analysis of beer, in: V.R. Preedy (Ed.), Beer in Health and Disease Prevention, Academic Press, San Diego, United States of America, 2009, pp. 943–962.
[34] Y. Ozaki, Applications in chemistry, in: H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), Near-Infrared Spectroscopy. Principles, Instruments, Applications, Wiley-VCH, Weinheim, Germany, 2002, pp. 179–212.
[35] M.J. Martelo-Vidal, M. Vázquez, Determination of polyphenolic compounds of red wines by UV-VIS-NIR spectroscopy and chemometrics tools, Food Chem. 158 (2014) 28–34.
[36] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal. 44 (3) (2007) 683–700.
[37] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Chapter 36: multivariate calibration, in: A. Muñoz-de-la-Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri (Eds.), Data Handling in Science and Technology/Handbook of Chemometrics and Qualimetrics: Part B, vol. 20, Elsevier, Amsterdam, Netherland, 1998, pp. 349–381.
[38] V. Sileoni, O. Marconi, G. Perretti, Near-infrared spectroscopy in the brewing industry, Crit. Rev. Food Sci. Nutr. 55 (12) (2015) 1171–1791.
[39] J.B. Reeves, G.W. McCarty, J.J. Meisinger, Near infrared reflectance spectroscopy for the determination of biological activity in agricultural soils, J. Near Infrared Spectrosc. 8 (3) (2000) 161–170.