Neural Model to Describe Microbial Concentration in the Bioreactor for Biosurfactant Production Using Waste Substrate

Caroline D. Grossi[a], Ana Maria F. Fileti[b], Brunno F. Santos[a]

[a]Department of Chemical and Materials Engineering, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea. Rio de Janeiro, RJ 22453-900, Brazil.
[b]School of Chemical Engineering (FEQ), Department of Chemical Systems Engineering (DESQ), University of Campinas (UNICAMP). Rua Albert Einstein, 500 – Cidade Universitária, Campinas – SP, 13083-852, Brazil.
bsantos@puc-rio.br

Biosurfactants are produced through metabolism of microorganisms (bacteria, yeast and fungi) and many applications are attributed to them. Biomass concentration is an important variable in biosurfactant production process, because is related to substrate consumption and production rate. This variable is collected by sampling and determined by off-line analysis with significant time delay, in most of cases. The present work was carried out the development of models based on artificial intelligence to predict biomass concentration faster than usual analytical method (dry weight). The process was performed in batch-bioreactor using waste substrate. Feedforward neural networks were compared to determine the best model due to the database set acquired from the bioreactor plant. Software MATLAB 2016b was used to implement artificial neural network. The available input layers were agitation, aeration, absorbance, glucose concentration, dissolved oxygen concentration, surface tension and surface tension dissolved in 10 × and 100 ×. The network topology was determined by the combination of parameters such as number of neurons, training algorithm and activation functions. The results showed that the models are appropriate to predict the biomass concentration profile with good agreement of R^2 (0.988), sum squared error (SSE) was 0.081 and mean squared error (MSE) was 0.0001.

1. Introduction

Microbial biomass consists mostly of bacteria and/or fungi which produce something of biological value. The microbial biomass is affected by several factors from industrial processes such as aeration, dissolved oxygen, pH and others, fact that proves a measurement is needed. Methods have been assessed for estimating biomass faster than traditional methodologies (as gravimetric analysis).

The biosurfactants are amphipathic molecules with tensoactive activity, able to reduce the interfacial tension of the medium in which they are dispersed. These molecules have been used in many industries, such as pharmaceuticals, food, oil, among others, acting as an ecological alternative to the use of synthetic surfactants. Interest in biosurfactant production has been increasing last years due to possible alternatives to chemical surfactants. Most of them are produced by fermentation of bacteria, fungi or yeasts (Nitschke and Pastore, 2002, Vaz et al., 2012). The biosurfactant production has been reported by Zheng et al. (2015), Wang et al. (2012), Sundaram and Thankur (2015) and others.

Different routes were performed to reduce biosurfactan production costs, even for increasing yields by the use of substrate based on waste (Secato et al. 2016, Santos et al., 2014). Also, the biosurfactants have been studied as a technological strategy to prevent unwanted problems caused by possible environmental accidents, as reposted by Rufino at al. (2016).

Monitoring of the variables of this process may result low-cost in biosurfactant production. In terms of industrial production, it is interesting to understand the behavior of the variable of interest through mathematical models, which may aid in the control of the process. Modeling techniques that use artificial

intelligence, such as modeling based on artificial neural networks (ANNs). They are advantageous alternatives since they allow faster monitoring of the variable that needs a long period of time to be measured.

ANN is a method similar to networks of interconnected biological neurons that was capable of handling multiple independent and dependent variables simultaneously and the functional relationship does not need to be known (Ahmad et al. 2017). The goal of ANN is to obtain a topology, which produces small errors between experimental data and predicted data through the process called training. The training adjusts weights and bias making the model learns the input-output behaviors of the process. When a ANN is capable to perform as well on data set the network generalizes well. The collected information is provided as input into the neural network model and the output variable is automatically estimated, most often being an unmeasured variable. When the number of connection weights is large, the ANN may not provide good generalization, called overfitting. It is possible to find several works that use the artificial intelligence methodology applied to the production of biosurfactants, as in Dhanarajan et al. (2014), Ahmad et al. (2016), Secato et al. (2017) and others.

In this work, a model based on neural networks was developed to predict the behavior of the microbial concentration in fermentation by *Bacillus subtilis*, using residues as substrate: glycerol from the biodiesel process and beet peel.

2. Methodology

2.1 Experiments

The experimental data in this work were studied by Santos (2015). The experiments were carried out in a 7L bioreactor (Bioflow 310 New Brunswick Scientific, USA). Each batch was performed for 24 h and the experimental data (used as neural model inputs) were collected every 3 h.

The microorganism for the production of biosurfactants was *Bacillus subtilis*, available from the Microorganism Bank of the Center for Research in Chemistry, Biology and Agriculture (CPQBA / Unicamp).

The substrate composition based on waste was 6% (v/v) glycerol from the biodiesel production and 7.5% (v/v) beet peel. Glucose concentration was inferred from a calibration curve, established by a laboratory biochemical test kit.

Microbial biomass in the process was determined by the dry weight method, where at the end of the tests a 30 ml sample of the culture broth was centrifuged (10000 rpm, 10 min, 40°C) and then dried at 50°C for 24h and the evaluated weight.

2.2 Artificial Neural Network Models (ANN's)

For the construction of the neural model, a study of the input variables was initially performed to evaluate which has the greatest influence on the process. In the sequence, the database was prepared, and divided into training data and test data in the ratio of 75% and 25%, respectively.

The variables were: agitation (AG - rpm), aeration (AE - vvm), absorbance (ABS), glucose concentration (CG - g/L), dissolved oxygen (DO - mg/L), surface tension (TS-mN/m) and surface tension diluted at 10 and 100x (TS1 and TS2 - mN/m). These variables were chosen because they are measured faster during the fermentation process, and absorbance is a qualitative measure of cell density, being directly linked to the variable one wishes to predict.

The variable of interest, chosen as output of the neural model, was the microbial concentration (CM-mg / L) due to a long period of time for its analytical determination, about 48h by the dry weight method previously described.

The MATLAB R2016b software was used as a support tool for modeling the system. Several combinations of ANN parameters were tested in order to find the best neural model capable of efficiently predicting the described process.

The training was developed using some algorithms to optimize parameters (weights and bias). It is used a gradient descent approach according to the negative of the error gradient. Approach based on Quasi-Newton method to adjust the parameters were tested such as the Levenberg Marquardt algorithm in conjunction with Bayesian regularization (trainbr).

Different Activation functions were used in the hidden layers: logistic sigmoidal (*logsig*), seen in Equation 1 and hyperbolic tangent (*tansig*), seen in Equation 2. In the output layer was used the linear activation function (*purelin*) Equation 3.

$$logsig = \frac{1}{(1+\exp(-n))} \tag{1}$$

$$tansig = \frac{2}{(1+\exp(-2n))} - 1 \tag{2}$$

$$purelin = n \tag{3}$$

Where *n* is processed signal by artificial neural networks.

The choice of the best neural model took into account the best values of SSE (Equation 4), MSE (Equation 5) and R^2 (Equation 6) for the data from the training data set.

$$SSE = \sum_{i=1}^{n}(Y_{observed} - Y_{predicted})^2 \tag{4}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_{observed} - Y_{predicted}) \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{observed} - Y_{predicted})^2}{\sum_{i=1}^{n}(Y_{observed} - Y_{mean})^2} \tag{6}$$

3. Results

The experimental data were interpolated using the cubic spline in order to increase the number of data, totalling 600 vectors. It is important to use the normalization factor for both training data and test data to ensure matching at the end of prediction.

Table 1 shows the adopted strategies for using variables as input layer. Thus, the analyses were performed by dividing into five blocks, with one and two intermediate layers.

Within each block, several scenarios were simulated by varying the number of neurons in input layer, the activation functions, number of neurons of the intermediate layers and the training algorithm in order to find the best neural model by block. Finally, a comparison was made between the best results of each block, and then the best neural model was determined capable of efficiently predicting the described process.

Table 1: Input variables divided by blocks.

Variables	Blocks				
	1	2	3	4	5
Glucose Concentration	X	X	X	x	X
Dissolved Oxygen	X	X	X	X	X
Absorbance	X	X	X	X	X
Agitation	X	X	X	X	X
Aeration		X	X	X	X
Surface Tension			X	X	X
Surface Tension Diluted at 10x				X	X
Surface Tension Diluted at 100x					X

The results of the all blocks are shown on Table 2, the topologies and performance values were assessed. The following scenarios is only a bit of all tested combinations.

All selected conditions in Table 2 showed high values of R^2 (>0.9) and error index (SSE and MSE) values very different. The best simulated scenario, comparing the results of all the blocks, represents the topology of 8x9x13x1 (in relation to the number of neurons of each layer), using the *logsig, logsig* in hidden layer. This topology showed an R^2 value for the test of 0.988, which means that the network has good predictability, since the test data was not previously presented to the network. The 98.8% fit found is considered very good for a neural network model. R2 higher than 90% indicated excellent agreement of the neural network model with the experimental training and validation values, obtained from database.

The Figure 1 shows the ANN topology in training box (Matlab) with 8-9-13-1, *logsig-logsig, purelin*, trained by *trainbr* resulting in 225 parameters (weight and bias).

For the latter case, an SSE error value of 0.08 was found, considered the lowest error of all the simulated scenarios, which shows precision in the prediction. The Figure 2 below illustrates the results found for the best scenario with dispersion graph and comparison of the experimental values (observed) and predicted values (calculated) with neural model.

Table 2: Scenarios of different topologies divided by blocks, with results of the performance R^2, SSE and MSE.

| Neurons (hidden layer 1)	Activation Function	Neurons (hidden layer 2)	Activation Function	R^2	SSE	MSE
Block 1						
15	logsig	-	-	0.946	3.02	0.00671
15	logsig	12	tansig	0.938	2.97	0.00660
Block 2						
9	logsig	-	-	0.960	1.28	0.00284
10	logsig	11	logsig	0.928	0.49	0.00108
Block 3						
15	logsig	-	-	0.977	0.99	0.00220
7	logsig	9	logsig	0.979	0.84	0.00186
Block 4						
20	tansig	-	-	0.961	0.70	0.00155
10	logsig	10	logsig	0.982	0.22	0,00048
Block 5						
14	logsig	-	-	0.981	0.74	0.00164
9	logsig	13	logsig	0.988	0.08	0.00017
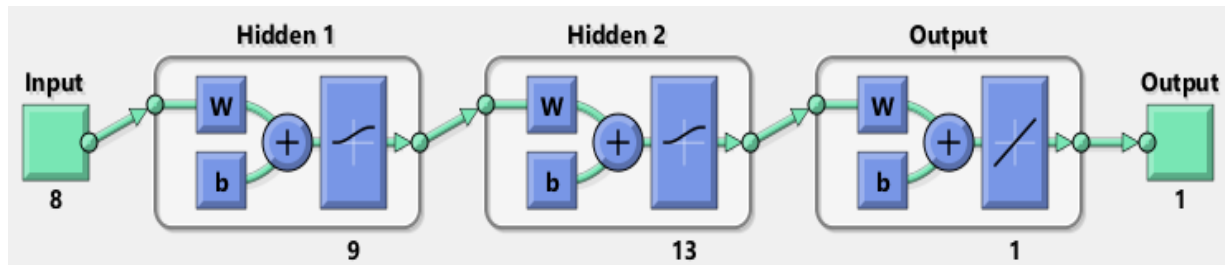
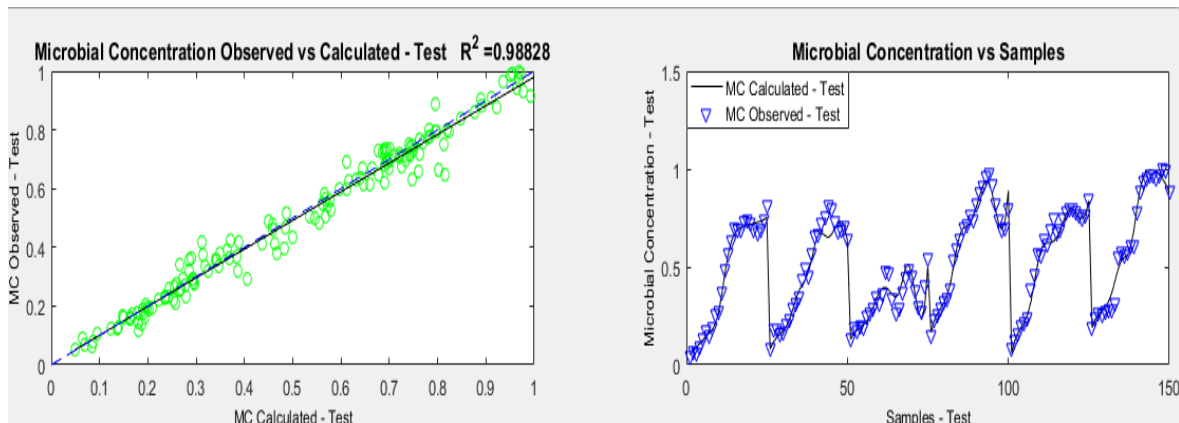*Figure 1. Neural network model developed.*

Figure 2. Observed graphs showing the best scenarios simulated for the process.

Combinations of the different input variables may affect the microbial concentration as shown in the different scenarios of the biosurfactant production process due to the complex relationship between system conditions. Neural model applied to estimate microbial concentration allowed to reach goal and the SSE and MSE indicated that the model well learned the relationship among the input and output variables.

Microbial concentration is an important variable in the biosurfactant production, but it takes long time to be carried out, more than 24 hours. Albuquerque et al. (2008) developed a neural model to estimate biomass of *Candida lipolytica* online and they used pH and dissolved oxygen as input variables, reaching R^2 of the 0.96 and error indexes SSE and MSE of the 0.09 and 0.0004, respectively.

Neural models for predicting biosurfactant production from waste is reported by Tayyebi et al. (2013). They used soybean oil refinery waste. This work proposed a topology for batch fermentation experiments of 10 neurons in the input layer, 8 neurons in the hidden layer and 4 (residual oils) in the output layer with R^2 of 0.971. The analysis evidenced that the residual oil has high effect on the biosurfactant production.

4. Conclusions

In this study, predicted dynamic models based on the neural network have been proposed for the microbial concentration in the biosurfactant production by *Bacillus subtilis* from waste. Topology was 8-9-13-1 with appropriate for R^2 and error indexes. The model was able to measure microbial concentration (g/L) properly, much less time than required through analytical methods. The blocking was a strategy that allowed the best comparison of the results, leading to the best topology of simulation.

The ANN development was performed offline, which means that during the training and the test were used data collected from experiments already performed. From the network training, one can build a soft sensor, capable of receiving the data collected during a new experiment, and carry out the online prediction. This can be very advantageous from the point of view of process control and optimization.

As a proposal for future work, comparisons between other types of ANN models, such as Elman's recursive networks, Hopfield, or other artificial intelligence techniques, such as a fuzzy logic study and neuro-fuzzy could be performed.

Reference

Ahmad, Z., Bahadori, A., Zhang, J., 2017, Prediction of combustion efficiency using multiple neural networks. Chemical Engineering Transactions, 56, 85-90, DOI: 10.3303/CET1756015.

Ahmad, Z., Crwley, D., Marina, N., Jha, S. K., 2016, Estimation of biosurfactant yield produced by *Klebseilla sp* FKOD36 bacteria using artificial neural network approach, Measurement, 81, 163-173.

Albuquerque, C. D. C., Takaki, G. M. C., Fileti A. M. F., 2008, On-line biomass estimation in biosurfactant production process by Candida lipolytica UCP 988, 35, 1425-1433, DOI: 10.1007/s10295-008-0443-5.

Dhanarajan, G.; Mandal, M.; Sem, R., 2014, A combined artificial neural network modeling-particle swarm optimization strategy for improved production of marine bacterial lipopeptide from food waste. Biochemical Engineering Journal, 84, 59-65.

Nitschke M.; Pastore, G. M., 2002, Biosurfactants: properties e aplications. Quím. Nova, 25,772-776, DOI:.10.1590/S0100-40422002000500013.

Rufino R.D., Neves da Motta Silveira G., Luna J.M., Sarubbo L.A., 2016, Conservation of the biosurfactant produced by pseudomonas aeruginosa for environmental applications, Chemical Engineering Transactions, 49, 535-540 DOI: 10.3303/CET1649090

Santos B., Ponezi A., Fileti A.M.F., 2014, Strategy of using waste for biosurfactant production through fermentation by bacillus subtilis, Chemical Engineering Transactions, 37, 727-732 DOI: 10.3303/CET1437122

Santos B.F., 2015, Study of biosurfactant production using agro-industrial waste with development of statistical models and soft sensor by artificial neural network. Doctoral Thesis– University of Campinas. Campinas/SP-Brazil.

Secato, J. F. F., Santos, B. F., Ponezi, A. N., Tambourgi, E. B., 2017, Optimization techniques and development of neural models applied in biosurfactant production by Bacillus subtilis using alternatives substrates, Advances in Bioscience and Biotechnology, 8, 343-360, DOI: 10.4236/abb.2017.810025.

Secato J., Coelho D., Rosa N., Lima L., Tambourgi E.B., 2016, Biosurfactant production using bacillus subtilis and industrial waste as substrate, Chemical Engineering Transactions, 49, 103-108 DOI: 10.3303/CET1649018

Sudaram, S., Thankur, S., 2015, Biosurfactant production by a CO2 sequestering *Bacillus sp.* strain ISTS2, Bioresour. Technol., 188, 247-250, DOI: 10.1016/j.biortech.2015.01.029.

Tayyebi, S., Lotfabad, T. B., Roostaazad, R., 2013, Applying neural network to dynamic modelling of biosurfactant production using soybean oil refinery wastes, Iranica Journal of Energy & Environment, 4, 161-170, DOI: 10.5829/idosi.ijee.2013.04.02.14.

Vaz, D.A.; Gudiña, E.J.; Alameda, E.J.; Teixeira, J.A; Rodrigues, L.R., 2012, Performance of a biosurfactant produced by a *Bacillus subtilis* strain isolated from crude oil samples as compared to commercial chemical surfactants. Colloids and Surfaces B: Biointerfaces, 89, 167–174.

Wang, X. B., Nie, Y., Tang, Y. Q., Wu, G., Wu, X. L., 2012, n-Alkane chain length alters *Dietzia sp.* strain DQ12-45-1b biosurfactant production end surface activity, Appl. Environ. Microbiol., 79, 400-402, DOI:10.1128/AEM.02497-12.

Zheng, T., Xu, Y. S., Yong, X. Y., Li, B., Yin, D., Cheng, Q. W., Yuan, H. R., Yong, Y. C., 2015, Endogenously enhanced biosurfactant production promotes electricity from microbial fuel cells, Bioresour. Technol., 197, 416-421, DOI: 10.1016/j.biortech.2015.08.136.