CHEMICAL ENGINEERING TRANSACTIONS

VOL. 32, 2013

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at: www.aidic.it/cet

Chief Editors: Sauro Pierucci, Jiří J. Klemeš

ISBN 978-88-95608-23-5; **ISSN** 1974-9791

DOI: 10.3303/CET1332158

Development of a Monitoring Hybrid System for Bioethanol Production

William E. Herrera*, Rubens Maciel Filho

School of Chemical Engineering, State University of Campinas, P.O. Box 6066, 13083-970 Campinas, SP, Brazil
williamweha@hotmail.com

In different parts of the world researches are focusing in the development of new technologies for liquid fuel production based on renewable resources. In this context, as a major producer of sugarcane-based bioethanol, Brazil aspires to reduce its fossil fuel consumption and its associated impacts on environment. These types of products are becoming an important option for the country sustainable development. In this context the broad objective of this work is to propose and evaluate an efficient way of reducing bioethanol fuel production costs, through the development of tools that allow the process to run under control, even when possible fluctuation on feedstock are present.
In this work it was developed a software sensor to infer the concentration rates of substrate, biomass and product from secondary measurements of pH, turbidity, CO_2 flow rate and temperature. The software sensor uses a hybrid neural model to combine a multi-layer Artificial Neural Network (ANN) and the mass balance which describes the fermentation process kinetics. The experimental data used on the hybrid model training were obtained from fermentations that took place between 30 ºC and 38 ºC, measured at every 2 ºC interval. The raw material for the fermentation is a mixture of 75 % hydrolyzed sugarcane bagasse and 25 % sugarcane molasses. This kind of composition may be a possible feedstock for is typical of second generation ethanol production. The three Hybrid Neural models developed are robust models that describe adequately the fermentation process even in the presence of changes in operating conditions. This is therefore a powerful tool for the prediction of kinetic rates in fermentation processes, which eventually may be used for online applications.

1. Introduction

Among the current features that have to be improved to increase the bioethanol process performance is that related to the robustness of the alcoholic fermentation when quality fluctuations in the raw material occur, which allow changing in the kinetic behavior that have influence on the yield, productivity and conversion of the substrate in the main final product. Type and quality of the sugarcane, harvest time and the sucrose content are among the factors that could interfere in the fermentation process. This interference can be compensate with adjustments in operation conditions and process control. Therefore, it is necessary to have an efficient monitoring system using reliable sensors in order to keep the fermentation process at the desired conditions, avoiding or minimizing the disturbance influence due raw material quality variations as well as microorganisms metabolic alterations. As in most bioprocesses, the key variables in the alcoholic fermentations are the concentrations. Online concentration measurements are usually expensive and might delay the process significantly. It should also be considered the complexity associated with fermentation kinetics, nonlinearities and other characteristics of this process type that brings difficulties in the control of biological processes. The nonlinearities in the growth and production phases make the model development essentially difficult. Additionally, on developing operation and control strategies, it must take into account that in industrial environment there are difficulties in implementing, in a cheap, precise and robust way, devices that could be used in online measurements of substrate concentrations, biomass and products of interest. (Srivastava and Gupta, 2011), (Hugan, 2012). Following chemical industry example, it is necessary to study measurement ways that focus on precision and economic viability to facilitate evaluation, development and implementation of optimization and control

Please cite this article as: Herrera W., Maciel Filho R., 2013, Development of a monitoring hybryd system for bioethanol production, Chemical Engineering Transactions, 32, 943-948 DOI: 10.3303/CET1332158

techniques (Arauzo Bravo et al, 2004). The software sensors are algorithms used to infer variables that are difficult or expensive to be measured through the use of variables that are easy and cheap to be measured (secondary variables). In the present case study, the Artificial Neural Network (ANN) is used as a software sensor. This approach can be a suitable alternative to online measurements of concentration and controllers as it is shown in the works of Chen et al.(2004); Hocalar et al.(2011); Herrera (2012), who used ANN for concentration predictions. Preliminary results from Radke (2002) and Herrera (2012) show that it is possible to use simple secondary measurements like pH, turbidity, CO_2 flow rate, and temperature to infer biomass, substrate and product concentrations in alcoholic fermentation processes, all of them having strong correlation to the experimental data. RNA algorithms were tested as a control tool in biotechnological processes; preliminary results from (Andrásik et al., 2004; Srivastava et al., 2011) show that they can be used to make adequate control and monitoring.

The advantages shown by software sensors are used in this work to develop a kinetic model and a softsensor algorithm for the fermentation process that infers the kinetics of the first and second generation ethanol production using a solution of hydrolyzed sugarcane bagasse and sugarcane molasses in proportions of 75 % and 25 % respectively. The data obtained are used to develop a software sensor for determining cell growth rate (r_x), substrate consumption rate (r_s) and product formation rate (r_p), from secondary measurements such as pH, turbidity, CO_2 flow rate and temperature.

2. Experiments

The experimental data were provided by Andrade (2012). Type Bach fermentations were carried out for developing and validating the software sensor. They were performed varying the fermentation temperature from 30, 32, 34, 36 and 38 °C and keeping fix the initial substrate concentration at 180 Kg/m^3. *Saccharomyces cerevisiae* grown in the laboratory of Bioprocess Engineering, at Unicamp's Faculty of Food Engineering, were used in these experiments.

3. Hybrid Neural Model Based in ANN and Mass Balance

The software sensor is composed by a hybrid and adaptive neural model that combines mass balance equations with Artificial Neural Network (ANN) of the Multilayer Perception (MLP) type, which in turn describes the kinetics rates r_x, r_s, r_p.

3.1 Balance Based Modeling

The balance model based in phenomenological approach comprises the mass balance equations, microorganism growth, substrate consumption and ethanol formation for a Bach reactor described as follows.

$$\frac{dX}{dt} = r_x \qquad (1)$$

$$\frac{dS}{dt} = -r_s \qquad (2)$$

$$\frac{dP}{dt} = r_p \qquad (3)$$

where X is the cell mass concentration (kg/m^3), S is the substrate concentration (kg/m^3) and P is the ethanol concentration (kg/m^3).

Before using the equations, it is important to define units as the cell growth rates r_x (kg/m^3 h), substrate consumption rate r_s (kg/m^3 h), and product formation rate r_p (kg/m^3 h).

3.2 ANN Structure Selection

Rivera et al.(2010) shows that the continuous function which describes the dynamic behavior of an alcoholic fermentation can be approximated within any desired accuracy with a MLP neural network, which has a hidden layer, using a sigmoid activation function and a final layer of linear function neurons. This MLP network with three layers is displayed in Figure 1a: the first is the input layer where the data is inserted, the second is the hidden layer, and the third is the output layer. The second layer can comprise different numbers of neurons leading to different prediction performances. The quality of the prediction is evaluated in terms of Mean Square Error (MSE). In this layer the transfer function is the Log-Sigmoid

Transfer Function. The output layer contains only one neuron with a Linear Transfer Function. In order to develop the software sensor, it was used the program Matlab ®, in which these functions are available.

(a) (b)

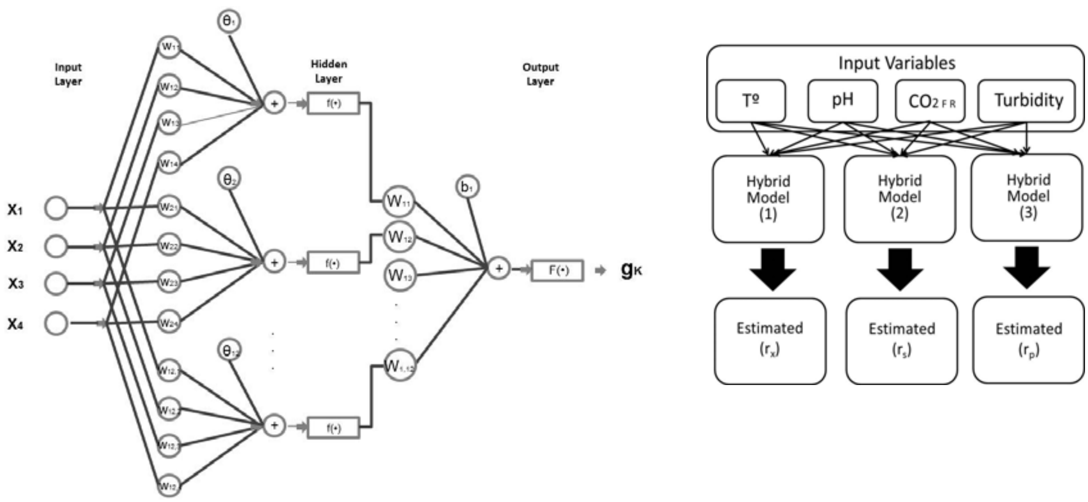

Figure1: a. General structure of the neural network. b. Software sensor framework based on the ANN

The mathematical relation is given by:

$$y_j = f\left(\sum_{i=1}^{N} w_{ij} x_i + \theta_j\right) \qquad (j = 1, \ldots.., M) \tag{4}$$

$$g_k = G\left(\sum_{i=1}^{N} W_{kj} x_j + \beta_k\right) \qquad (k = 1, \ldots.., K) \tag{5}$$

where w_{ji} is the weight of the connection of the *i*th neuron in the input layer and the *j*th neuron in the hidden layer; θ_j is the *j*th neuron bias in the hidden layer; W_{kj} is the weight of the connection between the *j*th neuron to the *k*th neuron in the output layer; β_k is the *k*th neuron bias in the output layer; F() and G() are the neurons activation functions in the hidden layer and output layer, respectively.

In this study, it was proposed three hybrid models for each of the outputs (kinetic rates) r_x, r_s, r_p. It is illustrated in Figure 2 the framework of the software sensor based on the developed ANN.

3.3. Selection of the Inputs Variables, Outputs Variables and Training

Preliminary results from Radke's (2002) reveal the possibility of using simple secondary measurements such as pH, turbidity, CO_2 flow rate, and temperature to infer biomass, substrate and product concentrations in an alcoholic fermentation. Based on this, these variables were selected as input data in Hybrid Model Neural proposed in this work. The data outputs of the ANN of the process are the kinetic rates r_x, r_s, r_p.

It was created a neural network for each variable under study to solve the kinetics rate problems (Figure 1 b). The death rate of the cells was not used to conduct this study since the viability has always been 100 % for all batches. The training process of the standard performance function for the neural network of the feedforward type is the mean square error between the input data and the desired objectives, defined in equation (6).

$$f = MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \tag{6}$$

For training, validating and testing the MLP network the Levenberg-Marquardt numerical optimization method is used in the standard performance function.

3.4. Selection of Historical Data

The input data for the network are provided by pH sensors, turbidimeter, CO_2 sensor and thermocouple to measure temperature, all of them with a sample rate of 3 minutes. Different ANNs were trained to describe each kinetic rate: cell growth rate (r_x), substrate consumption rate (r_s) and product formation rate (r_p). These are the output information in the ANNs.

In this work, the input data are available from direct measurement of the sensors every 3 minutes. The output data, that is, the kinetic rates that are determined from the concentrations of cell mass (X), substrate (S), product (P), are gathered by means of analytical measurements methods: HPLC for concentrations and gravimetry for cell mass (Andrade, 2012). These experimental output data, however, were made in a time interval of approximately 3 hours. In order to obtain a correspondence between input and output data it was adjusted a sigmoid equation to the experimental data of concentrations biomass, substrate and product over time. The derivative of equation (7), adjusted for X, S and P, provided the values of dX/dt, dS/dt and dP/dt. This value, substituted in equations (1-3), result in the so called "experimental" kinetic rates used as outputs in the training, validation and testing.

It was developed three sets of representative data, comprising a total 8451 input/output patterns and corresponding to approximately 137 hours of fermentation. From this database 7606 patterns were selected and presented to each ANN related to the kinetic rates r_x, r_s, r_p, at temperatures between 30, 32, 34, 36 and 38 °C, with fixed initial substrate concentration at 180 g/L. The validation was performed, i.e. the predictive ability of the neural network, in a different sequence from the one used in the learning process, which comprises the remaining 845 patterns in the database.

$$y = a + \frac{b}{1 - e^{-\left(\frac{c - time}{d}\right)}} \tag{7}$$

4. Result and Discussion

The model based on neural network has been enhanced by selecting the best network architecture considering the number of neurons in the hidden layer, evaluating the predictions quality for the data set. This work was carried out by evaluating the mean square error (MSE), the R value and the Residual Standard Deviation (RSD, %), as observed in Tables 1 to 3. The RSD is used to make a more accurate assessment of the predictive ability of the model. Deviations smaller than 10 % are considered acceptable in the bioprocess engineering (Atala et al.,2001; Rivera et al., 2007).

$$RSD = \frac{\left(\frac{1}{n}\sum_{k=1}^{n}(d_k - g_k)^2\right)^{0.5}}{\overline{d_k}} \times 100 \tag{8}$$

Based on these criteria the ANN for product formation rate (r_p) with 3-12 neurons in the hidden layer was evaluated, and the best performance was obtained at 10 neurons in the hidden layer. With this arrangement, the values of MSE= 0.0039, an RSD(%)= 3.7057 and R= 0.9976 provide the most accurate prediction, since they lead to the minimum MSE, a R value near 1 (which means a good adjustment of the ANN), and a RSD value smaller than 10 %, as can be seen in Table 1.

Table 1. Test in the ANN for product formation rate (r_p), for 30, 32, 34, 36 and 38º C

Neurons in the hidden layer	MSE	R	RSD (%)
3	0.363	0.734	24.934
5	0.012	0.992	5.205
6	0.036	0.976	8.088
8	0.004	0.998	3.826
10	**0.004**	**0.998**	**3.706**
12	0.003	0.998	7.652

Analyzing the ANN for substrate consumption rate with variations of 3-18 neurons in the hidden layer it is possible to concluded that the structure with 14 neurons employed in this layer with the values of MSE = 0.226, a RSD= 8.722 and a R value = 0.974 was the most accurate on prediction, since the MSE is minimum, the R value close to 1 and the value of RSD is lower than 10 %, as shown in Table 2.

Table 2. Test in the ANN for substrate consumption rate (r_S), for 30, 32, 34, 36 and 38° C.

Neurons in the hidden layer	MSE	R	RSD (%)
3	0.518	0.905	30.553
5	0.726	0.913	30.553
6	0.654	0.920	13.921
8	0.530	0.939	13.894
10	0.355	0.961	9.646
12	0.355	0.961	9.007
14	**0.226**	**0.974**	**8.722**
16	0.299	0.967	9.449
18	0.518	0.905	30.553

The ANN for cell growth rate had variations from 3 to 20 neurons in the hidden layer during learning phase. From this, it can be stated that the structure with 18 neurons with MSE = 0.0000284, RSD = 3.64 and R = 0.998, was the most accurate on prediction since MSE is minimal, R close to 1 and the RSD value is lower than 10%, as depicted in Table 3.

Table 3. Test in the ANN for cell growth rate (r_X), for 30, 32, 34, 36 and 38° C

Neurons in the hidden layer	MSE	R	RSD (%)
3	0.000536	0.973	16.154
5	0.000339	0.983	14.930
6	0.000099	0.995	7.097
8	0.000108	0.994	7.221
10	0.000096	0.995	6.568
12	0.000236	0.997	5.344
14	0.000047	0.997	4.847
16	0.000031	0.998	3.741
18	**0.000028**	**0.998**	**3.640**
20	0.000050	0.997	4.880

The results are shown in the Figure 2 and 3. It can be noted that the hybrid model follows the desired trajectory of the experimental observations for the rates of cell growth, substrate consumption and ethanol production at 34 and 38 ° C. These temperatureswere taken as reference.

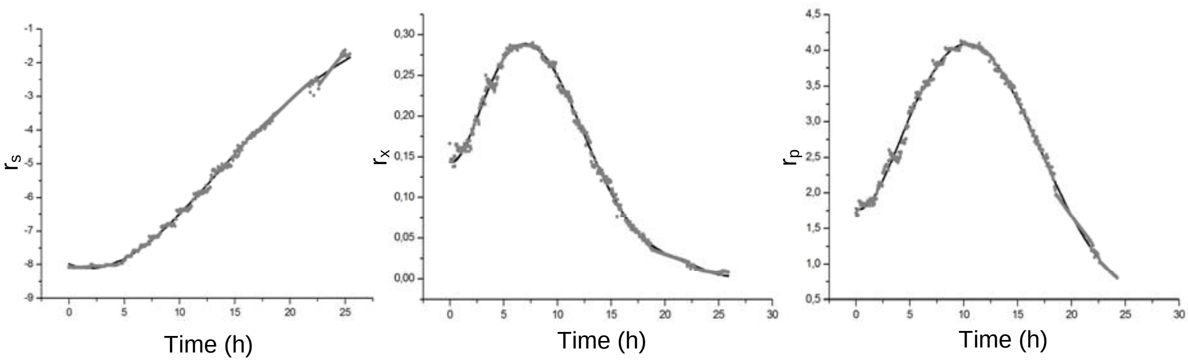

Figure 2. Experimental substrate consumption rate r_s, cell growth rate r_x, and product formation rate r_p (line) and hybrid neural model (filled circles), results at 34 °C.

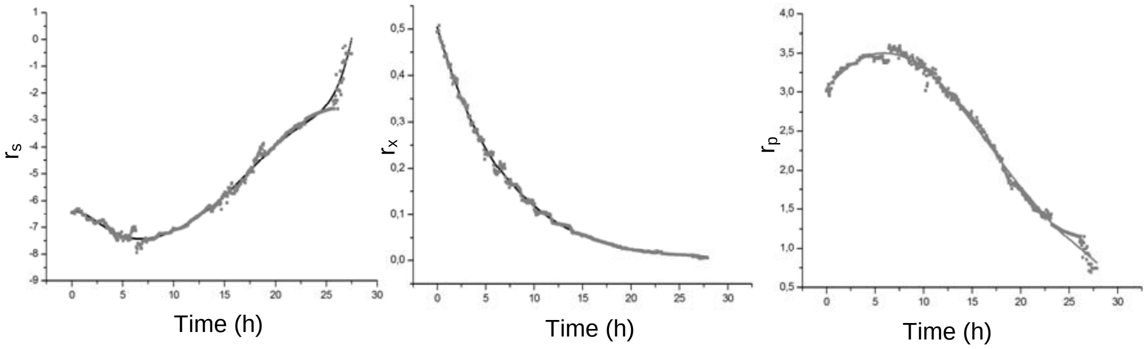

Figure 3. Experimental substrate consumption rate r_s, cell growth rate r_x, and product formation rate r_p (line) and hybrid neural model (filled circles), results at 38 °C.

5. Concluding Remark

This work showed that the developed hybrid model using as input secondary variables, which combines ANN and mass balances, can yield efficient results approximations in comparison with experimental data. This demonstrates the procedure ability to describe the fermentation process with high accuracy and precision.

The three Hybrid Neural models presented are robust models that describe adequately the fermentation process even in the presence of changes in operating conditions. This is therefore a powerful tool for the prediction of kinetic rates in fermentation processes. Variations in the kinetics rates are expected to occur in several situations, especially due to changes in the raw material quality or when fermentable sugar from hydrolysis is mixed with sugar cane molasses.

The developed models, based on easily measured variables, allow a better process monitoring since this method allows the implementation of an online estimation procedure.

References

Andrade, R. R., Costa, A.C., 2012, Kinetic of Modeling Ethanol Production Process from Enzymatic Hydrolysates of Sugarcane Bagasse Concentrated with Molasses Considering Cell Recycle. Campinas, São Paulo, Brazil: Chemical Engineering College, UNICAMP. (PhD Thesis in Portuguese).

Andrášik, A., A. Mészáros., de Azevedo. S. F, 2004, On-line Tuning of a Neural PID Controller Based on Plant Hybrid Modelling, Computers & Chemical Engineering 28: 1499-1509.

Arauzo Bravo, J. M., Cano Izquierdo, J.M, 2004, Automatization of a Penicillin Production Process with Soft Sensors and Adaptive Controller Based on Neuro Fuzzy Systems, Control Engineering Practice 12: 1073-1090.

Atala, D.I.P., Costa, A.C., 2001, Kinetics of Ethanol Fermentation with High Biomass Concentration Considering the Effect of Temperature, Applied Biochemistry and Biotechnology 91-93: 353-364.

Chen, L., Nguang, S. K., Xiao Dong, Chen., Xue Mei, Li., 2004, Modeling and Optimization of Fed-Batch Fermentation Processes Using Dynamic Neural Networks and Genetic Algorithms, Biochemical Engineering Journal 22(1): 51-61.

Herrera, W., 2012. Development and Implementation of a Software Sensor for Monitoring Online of Bioprocess. Campinas: Faculdade de Engenharia Química, UNICAMP. Brazil.

Hocalar, A., Türker, M., Karakuzu, C., Yüzgeç, U., 2011, Comparison of Different Estimation Techniques for Biomass Concentration in Large Scale Yeast Fermentation, ISA Transactions 50(2): 303-314.

Hugan, W.-H., Shieh, G. S., Wang, F-S., 2012, Optimization of Fed-Batch Fermentation Using Mixture of Sugars to Produce Ethanol, Journal of the Taiwan Institute of Chemical Engineers 43(1): 1-8.

Radke, E., 2002, Development of Hybrid-Neural Models for Alcohol Studies and assessment of techniques for process optimization. Campinas, São Paulo, Brazil: Chemical Engineering College, UNICAMP.(PhD Thesis in Portuguese)

Rivera, E. C., Costa, A.C., 2007, Development of Adaptive Modeling Techniques to Describe The Temperature-Dependent Kinetics of Biotechnological Processes, Biochemical Engineering Journal 36(2): 157-166.

Srivastava, A. K., Gupta, S., 2011, 2.38 - Fed-Batch Fermentation – Design Strategies. Comprehensive Biotechnology (Second Edition). M.-Y. Editor-in-Chief:Murray. Burlington, (Vermont, US), Academic Press: 515-526