



## Applying Neural Network to Dynamic Modeling of Biosurfactant Production Using Soybean Oil Refinery Wastes

<sup>1</sup>Shokoufe Tayyebi, <sup>2</sup>Tayebe Bagheri Lotfabad and <sup>1</sup>Reza Roostaazad

<sup>1</sup>Department of Chemical and Petroleum Engineering,  
Sharif University of Technology, Azadi Ave. Tehran, Iran  
<sup>2</sup>National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

(Received: October 20, 2012; Accepted in Revised Form: February 7, 2013)

**Abstract:** Biosurfactants are surface active compounds produced by various microorganisms. Production of biosurfactants via fermentation of immiscible wastes has the dual benefit of creating economic opportunities for manufacturers, while improving environmental health. A predictor system, recommended in such processes, must be scaled-up. Hence, four neural networks were developed for the dynamic modeling of the biosurfactant production kinetics, in presence of soybean oil or refinery wastes including acid oil, deodorizer distillate and soap stock. Each proposed feed forward neural network consists of three layers which are not fully connected. The input and output data for the training and validation of the neural network models were gathered from batch fermentation experiments. The proposed neural network models were evaluated by three statistical criteria ( $R^2$ , RMSE and SE). The typical regression analysis showed high correlation coefficients greater than 0.971, demonstrating that the neural network is an excellent estimator for prediction of biosurfactant production kinetic data in a two phase liquid-liquid batch fermentation system. In addition, sensitivity analysis indicates that residual oil has the significant effect (i.e. 49%) on the biosurfactant in the process.

**Key words:** Batch fermentation • Biosurfactant • Dynamic modeling • Neural network

### INTRODUCTION

Refinery vegetable oil wastes, the inevitable by-products of vegetable oil production, have become inexpensive and abundant sources of carbon. The worldwide surplus of these wastes has created environmental problems and the economical advantages of the vegetable oil industry have been greatly affected. Given the highly reduced form of carbon in vegetable oils and by-products, their conversion to biosurfactants could result in yields higher than those obtained with the use of common sugars [1-3]. Realizing this potential, however, would require the modeling of a fermentation process suitable for the conversion of oily compounds to biosurfactants. In fact, a perfect model is usually recommended for many processes, such as optimization, control, fault detection and diagnosis. One of the methods used for process

modeling is solving fundamental principles in biochemical engineering [4, 5].

In many processes, due to the strong interaction between process variables and the complexity of the dynamic system, process modeling becomes a complex task [6]. A black-box approach based on input and output data of the process and the relationship between them is another method for process modeling. A powerful tool associated with this approach is the neural network that requires no prior knowledge of the fermentation process [7]. Hur *et al.* [8] developed an artificial neural network to estimate the growth pattern of microorganisms during a fermentation process. In another study, neural network models were used to predict the concentration profile of a hydrochloric acid recovery process [9]. Koprinkova and his colleagues [10] proposed the neural network modeling approach for cultivation of a microorganism using chemostat is desired.

**Corresponding Author:** Tayebe Bagheri Lotfabad, National Institute of Genetic Engineering and Biotechnology (NIGEB), Shahrak-e Pajooresh, km 15, Tehran-karaj Highway, Tehran, Iran, P.O.Box:14965/161.  
Tel: +98 21 44580321, Fax: +98 21 44580399, E-mail: bagheril@nigeb.ac.ir

To our best of knowledge, a comprehensive model has not yet been presented for estimating the biosurfactant production kinetics in the presence of different immiscible carbon sources. Here, a dynamic neural network model was proposed to describe the biosurfactant production process in a two-phase liquid-liquid medium containing soybean oil or refinery wastes including acid oil, deodorizer distillate and soap stock.

## MATERIALS AND METHODS

### Fermentation Kinetic Study and Analytical Methods:

The microorganism, *Pseudomonas aeruginosa* MR01 (EU795302) previously isolated from crude oil was pre-cultured overnight and cultivated in a 250 ml shake flask [11]. SOM medium was used as the growth medium, which included 3 g/l NaNO<sub>3</sub>, 0.25 g/l KH<sub>2</sub>PO<sub>4</sub>, 0.25 g/l MgSO<sub>4</sub>·7H<sub>2</sub>O, 1 g/l yeast extract and one of following carbon sources at concentration of 6- 8% (v/v): soybean oil, acid oil, deodorizer distillate or soap stock from a soybean oil refinery process. After pH adjustment to 7, flasks were incubated at 30°C, with shaking at 200 rpm [3]. Chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA). Soybean oil and its by-products were obtained from the Behshahr Industrial Company (BIC), one of the vegetable oil refineries in Iran.

To model *P. aeruginosa* MR01 in different broths, each containing one of the above-mentioned carbon sources, broth samples (50 ml) were withdrawn at specified times. The number of viable bacterial cells per ml of each sample was reported as colony-forming units (CFU). Sample broths were centrifuged (4,618 ×g, 4 °C, 30 min) and pellets were washed and dried to determine the dry cell weight (DCW). Subsequently, the supernatant was used to analyze oil consumption using the n-hexane extraction method [12]. Briefly, a 50 ml sample of cell-free culture was mixed vigorously with n-hexane 1:1 (v/v). The mixture was then centrifuged (7,441×g, 4°C, 20 min) to separate the phases. The organic phase was transferred to a new tube and evaporated to determine the residual oil concentration. One half of the aqueous phase (25 ml) was then used for subsequent nitrate assessment, according to the method by Jagessar and Sooknundun [13], in which the ammonium hydroxide solution was replaced with 12 N potassium hydroxide. The second half of the aqueous phase (25 ml) was used to quantify the rhamnolipid concentration via the acid precipitation and solvent extraction method involving ethyl acetate [14].

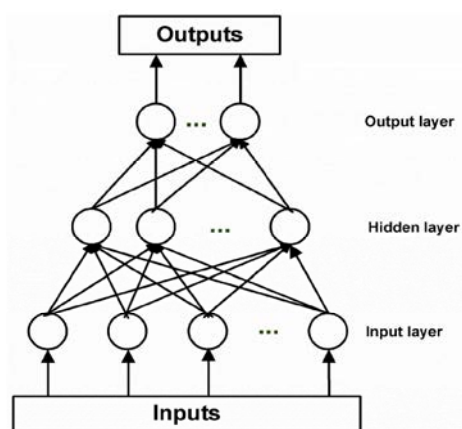


Fig. 1: Schematic of feed-forward network

**Neural Network Modeling:** A neural network consists of interconnected processing blocks of artificial neurons in which mathematical calculations are performed. The nonlinear computations are carried out by transfer functions in the neurons. The neurons are organized in layouts comprising one input layer, hidden layer (s) and one output layer. Each neuron is connected to every output from the preceding layer with weighted connections, according to the schematic of feed-forward network illustrated in Figure 1. During the training algorithm, the weights are adapted, based on the minimization of the mean square error (MSE), between the desired and the actual-network outputs [15].

Three statistical criteria, namely correlation coefficient ( $R^2$ ), root mean square error (RMSE) and standard error (SE) were used for assessing the performance of the developed neural network models used in this study. These are defined as:

$$R^2 = \frac{\sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2 - \sum_{i=1}^N (V_{\text{cal},i} - V_{\text{exp},i})^2}{\sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2} \quad (1)$$

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^N (V_{\text{cal},i} - V_{\text{exp},i})^2 \right]^{1/2} \quad (2)$$

$$SE = \frac{\sqrt{\sum_{i=1}^N (V_{\text{cal},i} - V_{\text{exp},i})^2}}{N-1} \quad (3)$$

where, N is the number of observations and  $V_{\text{calc},i}$  and  $V_{\text{exp},i}$  denote  $i^{\text{th}}$  output value of the neural network model and experimental data, respectively. Also,  $\bar{V}_{\text{exp}}$  is the mean

value of the experimental data. The small values of the RMSE and SE and the large value of the R<sup>2</sup> indicate that the neural network model is trained well.

For the dynamic modeling process, the feed-forward neural network model is expressed as:

$$y(t) = \psi \left( \begin{matrix} y(t-1), y(t-2), \dots, y(t-n), \\ u(t-1), u(t-2), \dots, u(t-m) \end{matrix} \right) \quad (4)$$

Where  $\psi$  is an unknown nonlinear function which must be approximated by the neural network,  $t$  is the time,  $u$  and  $y$  are the inputs and outputs of the process, respectively and  $n$  and  $m$  are the orders of the dynamic model for the output and the input, respectively.

The output and the input orders of the dynamic model for biosurfactant production under the above defined conditions are 2; therefore, the relationships between the outputs and their respective input variables are defined as:

$$BS(t) = \psi_1 (BS(t-1), BS(t-2), Res\_Oil(t-1), Res\_Oli(t-2)) \quad (5)$$

$$DCW(t) = \psi_2 \left( \begin{matrix} DCW(t-1), DCW(t-2), \\ Res\_Oil(t-1), Res\_Oli(t-2) \end{matrix} \right) \quad (6)$$

$$FCU(t) = \psi_3 \left( \begin{matrix} FCU(t-1), FCU(t-2), \\ Res\_Oil(t-1), Res\_Oli(t-2) \end{matrix} \right) \quad (7)$$

$$Nit(t) = \psi_4 \left( \begin{matrix} Nit(t-1), Nit(t-2), \\ Res\_Oil(t-1), Res\_Oli(t-2) \end{matrix} \right) \quad (8)$$

In equations 5-8,  $\psi_1$  to  $\psi_4$  are nonlinear functions and

$Res\_Oil$ ,  $BS$ ,  $DCW$ ,  $CFU$  and  $Nit$  indicate residual oil, biosurfactant, dry cell weight, colony forming unit and nitrate, respectively.

As shown in equations 5-8, for each output variable, its corresponding input variable is different unless there are two input variables (i.e.  $Res\_oil(t-1)$  and  $Res\_Oil(t-2)$ ). Therefore, a neural network should be developed for each output variable if a fully connected neural network structure is selected for the prediction of the process model. On the other hand, four neural networks should be developed for each substrate that can not be suitable. In order to overcome the aforementioned problem, a partially connected feed-forward network consisting of an input layer, a hidden layer and an output layer was selected for this study. The main advantage of

the proposed neural network is that one neural network should be trained for all outputs of the process, for each substrate. The numbers of neurons in the input and output layers are equal to 10 and 4, respectively.

The number of training data and inputs and outputs of the neural network can significantly affect the number of hidden neurons. Moreover, it is known that the number of hidden neurons is crucial for successful network training. Although the large number of hidden neurons may facilitate the training of the neural network, it increases the computational load and memory space rapidly. On the other hand, a small number of hidden neurons make the neural network extremely difficult or even impossible to be trained. Considering the above mentioned aspects, reaching a balance between the number of hidden neurons and the desired performance of the neural network, is a matter of great importance. Hence, in accordance to the neural network model proposed for the modeling of the present process, the optimum number of hidden neurons estimated by the trial-and-error method is 8.

Neurons between adjacent layers, defining the neural network structure (Figure 2), are not fully connected due to the fact that each of outputs is affected by some specific input (not all of them). For instance, the first output (i.e.  $BS(t)$ ) of the proposed neural network is a function of four inputs, as presented in equation 5.

A hyperbolic tangent sigmoid transfer function with the following expression is applied to the hidden and output layers:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (9)$$

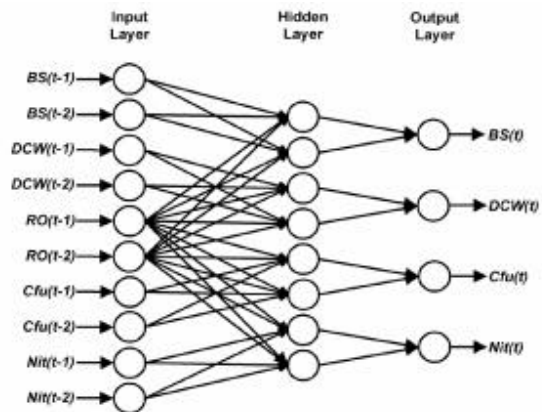


Fig. 2: Structure of the proposed neural network model for the biosurfactant production from soybean oil refinery wastes

In order to normalize the input and output between 0 and 1, using equation 10 is thus essential to match the proposed transfer function.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (10)$$

where  $x_{norm}$ ,  $x_{min}$  and  $x_{max}$  are normalized, minimum and maximum values of variable  $x$ , respectively.

### RESULTS AND DISCUSSION

The neural network used in this research was developed by MATLAB through the neural network toolbox. The total number of the experimental data was 100, of which 75% was used for training and the validation set and 25% was considered as the testing data. The neural network was properly trained by using the gradient descent method.

Variation in MSE with the number of iterations (epochs) for training and validation of the assumed neural network is presented in Figure 3. To prevent overfitting; the performance of the neural network was evaluated through the validation data after each iteration of the neural network training process. When the MSE of the validation set begins to increase, the training is stopped

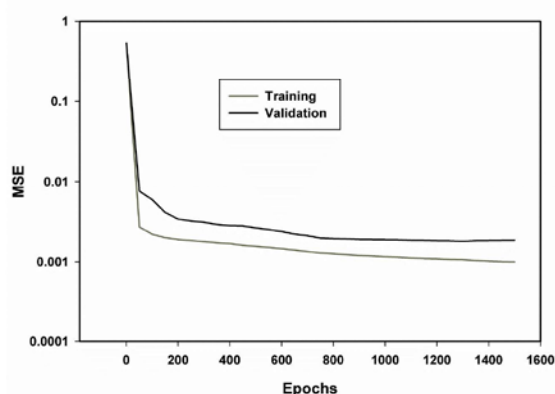


Fig. 3: Performance error is based on the training and validation epochs

[16]. Figure 3 shows that the individual condition is satisfied around iteration 1200.

Data predicted by the neural network and experiments with respect to time, are compared in Figures 4(A), 4(B), 4(C) and 4(D) and correlated by regression plots in Figures 5(A), 5(B), 5(C) and 5(D). The presented results in Figures 4 and 5 reveal the perfect model matching between the predicted neural network values and the experimental data.

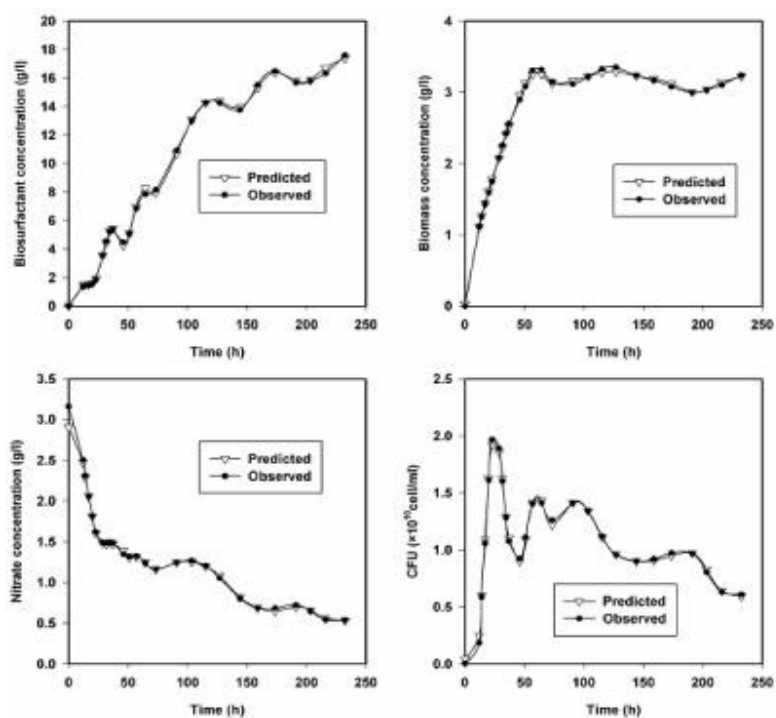


Fig. 4(A): Comparison between predicted neural network data and experimental data versus time in the presence of soybean oil

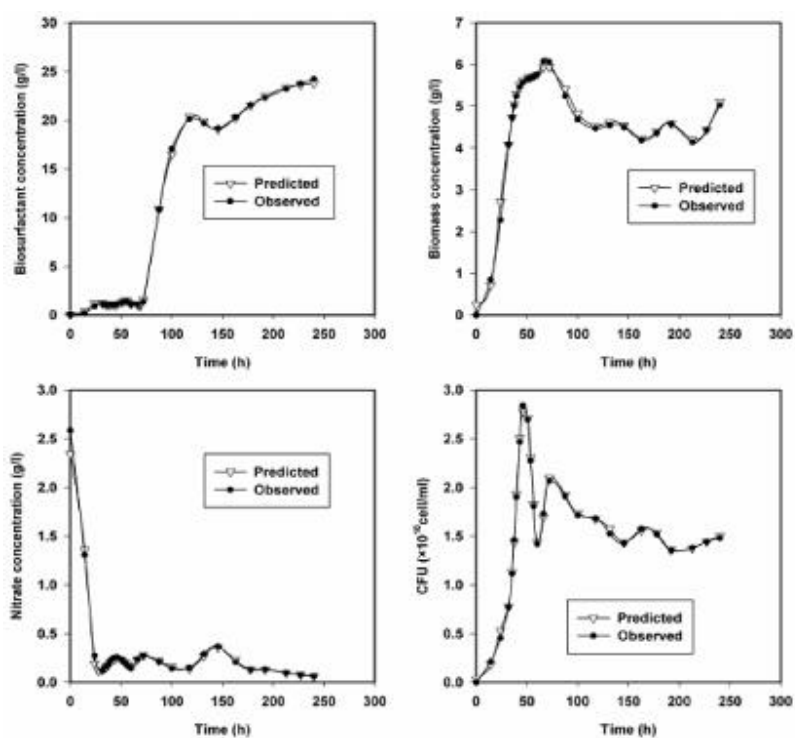


Fig. 4(B): Comparison between predicted neural network data and experimental data versus time in the presence of acid oil

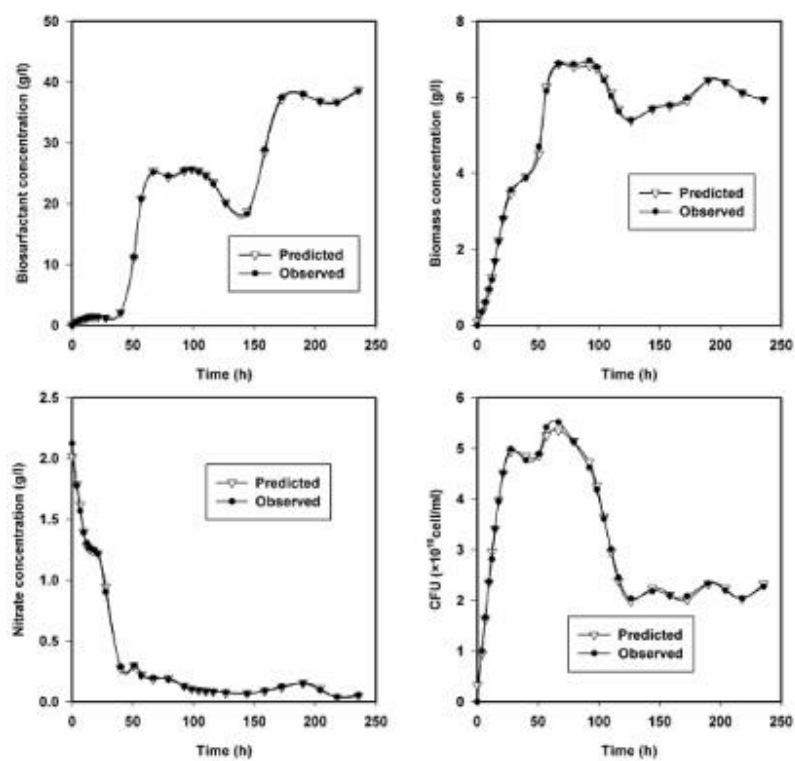


Fig. 4(C): Comparison between predicted neural network data and experimental data versus time in the presence of deodorizer

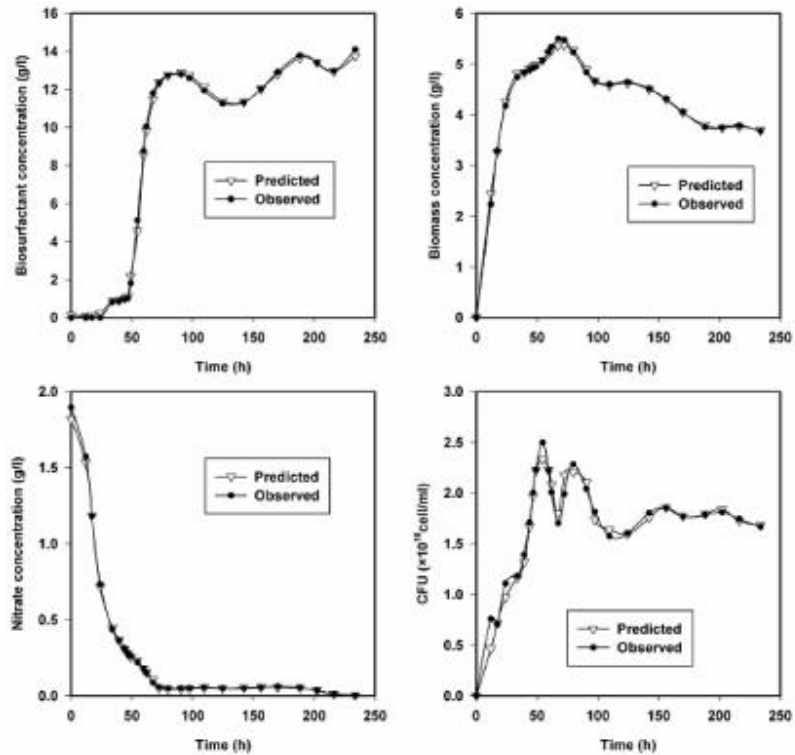


Fig. 4(D): Comparison between predicted neural network data and experimental data versus time in the presence of soap stock

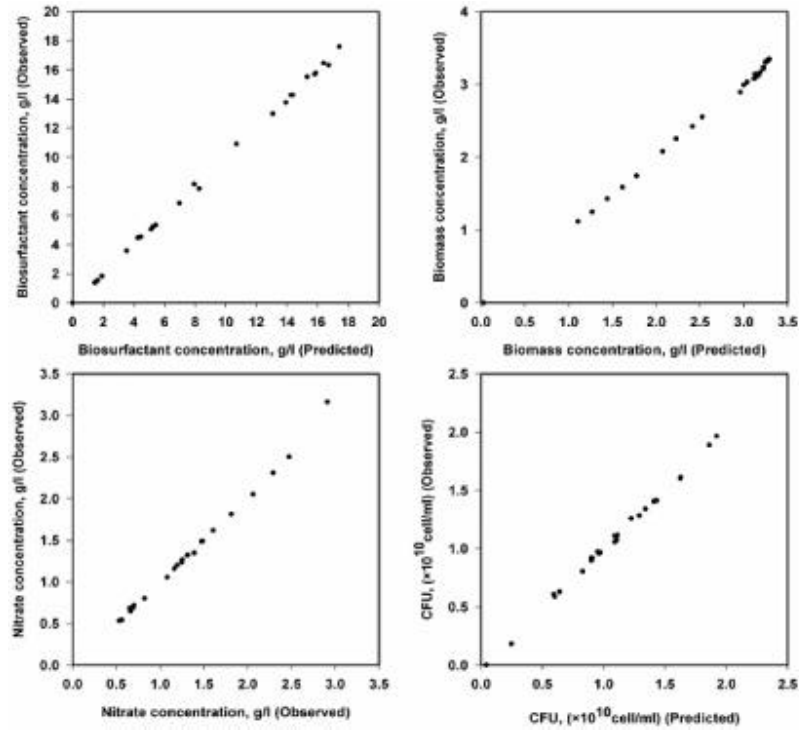


Fig. 5(A): Regression analysis between predicted neural network values and experimental data in the presence of soybean oil

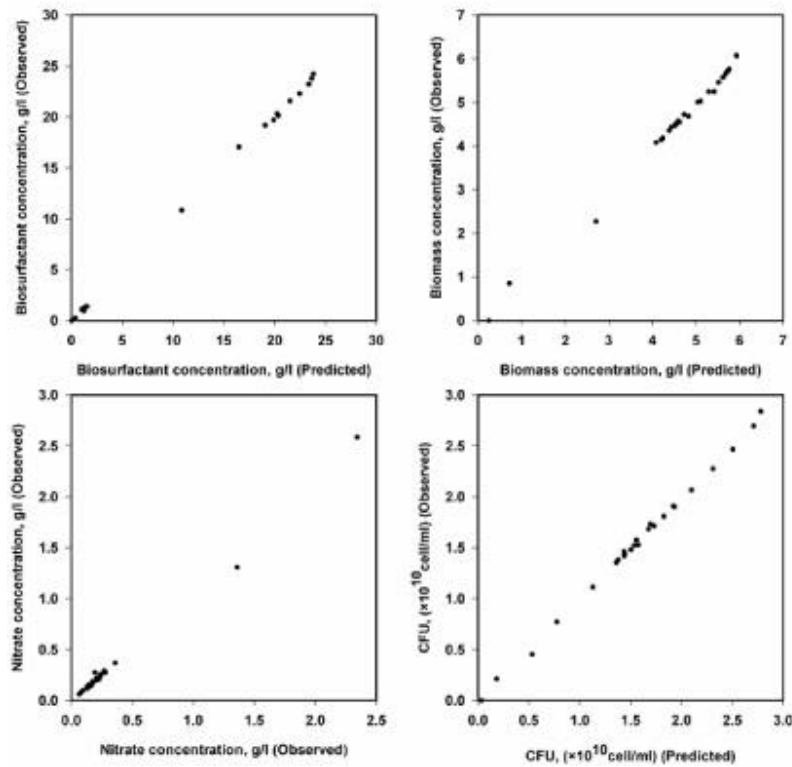


Fig. 5(B): Regression analysis between predicted neural network values and experimental data in the presence of acid oil

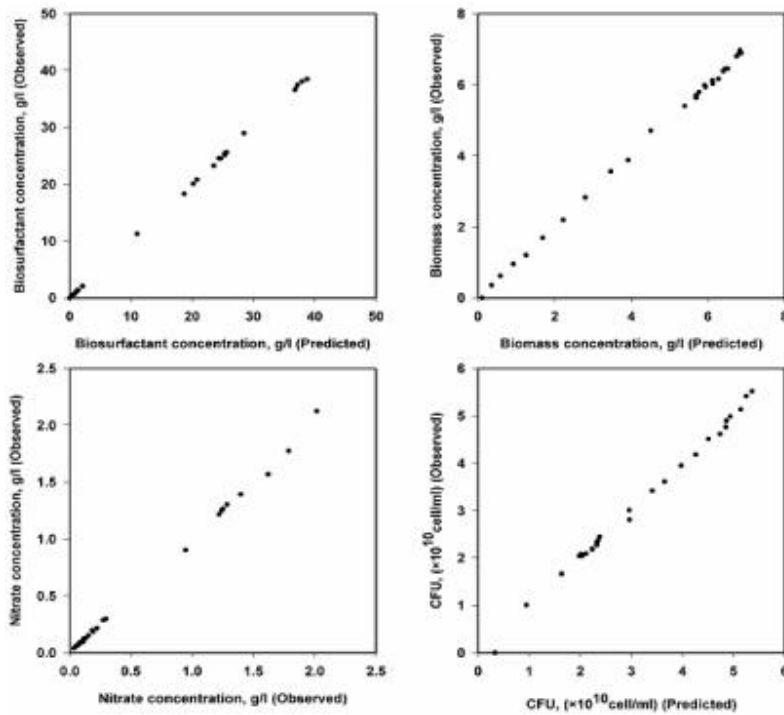


Fig. 5(C): Regression analysis between predicted neural network values and experimental data in the presence of deodorizer

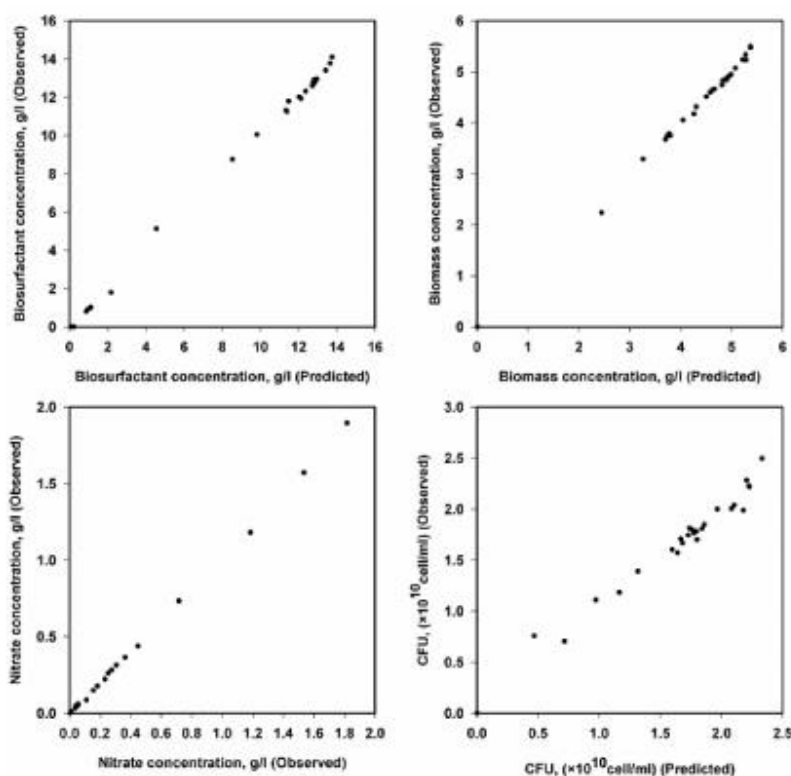


Fig. 5(D): Regression analysis between predicted neural network values and experimental data in the presence of soap stock

Table 1: Statistical analysis between predicted neural network data and experimental data.

| Output  | Soybean oil    |       |       | Acid oil       |       |       | Deodorizer     |       |       | Soap Stock     |       |       |
|---------|----------------|-------|-------|----------------|-------|-------|----------------|-------|-------|----------------|-------|-------|
|         | R <sup>2</sup> | RMSE  | SE    | R <sup>2</sup> | RMSE  | SE    | R <sup>2</sup> | RMSE  | SE    | R <sup>2</sup> | RMSE  | SE    |
| BS      | 0.999          | 0.161 | 0.033 | 0.999          | 0.183 | 0.037 | 0.999          | 0.185 | 0.028 | 0.999          | 0.198 | 0.041 |
| DCW     | 0.998          | 0.034 | 0.007 | 0.993          | 0.123 | 0.025 | 0.999          | 0.069 | 0.014 | 0.997          | 0.062 | 0.013 |
| CFU     | 0.997          | 0.025 | 0.005 | 0.998          | 0.029 | 0.006 | 0.996          | 0.095 | 0.010 | 0.971          | 0.091 | 0.019 |
| Nitrate | 0.993          | 0.053 | 0.011 | 0.989          | 0.052 | 0.011 | 0.998          | 0.026 | 0.005 | 0.999          | 0.019 | 0.004 |

The performance of the proposed neural network model was evaluated by statistical parameters (Table 1), where an R<sup>2</sup> value greater than 0.971 and RMSE and SE values less than 0.198 and 0.041, respectively, indicate the success of the neural network developed for modeling the production of biosurfactant by fermentation of soybean oil and wastes.

In this study, in order to evaluate the effectiveness of the input variables on the output variables, sensitivity analysis was also conducted. Accordingly, the sensitivity of each input is defined as:

$$Sensitivity = \frac{\%change\ in\ output}{\%change\ in\ input} * 100 \quad (11)$$

During the analysis of sensitivity, an input was varied between the mean value ± one standard deviation, whereas other inputs were held constant at their mean value. The sensitivity of the input variables on the output variables based on the experimental data and the developed neural network model is illustrated as a graphical bar in Figure 6. As shown in this figure, the sensitivity results of the developed neural network model are similar to the experimental data, with the percent error of less than 4%. In addition, the residual oil with average sensitivity of 49% has the significant effect on the biosurfactant in comparison with other outputs. Comparing the sensitivities showed higher sensitivity of biosurfactant concentration to residual oil concentration



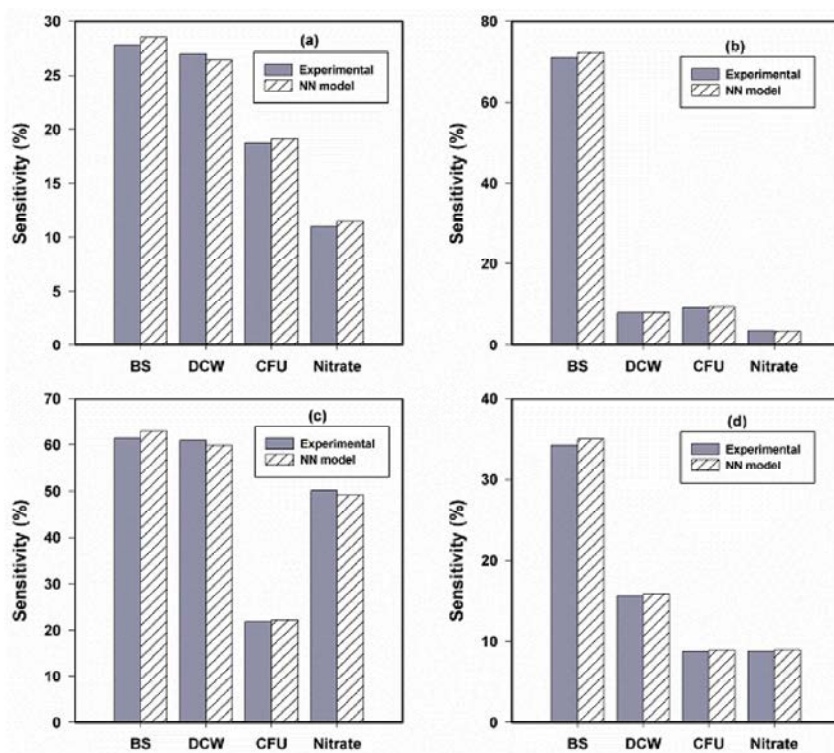


Fig. 6: Sensitivity analysis based on the proposed neural network model and the experimental data in the presence of (a) soybean oil, (b) acid oil, (c) deodorizer and (d) soap stock. (BS, biosurfactant; DCW, dry cell weight; CFU, colony forming unit)

which may concern presence of large number of fatty acids in biosurfactant molecule structure. In other words, this phenomenon seems to be originated from chemistry of biosurfactant.

### CONCLUSION

In this paper, predicted dynamic models based on the neural network have been proposed for the production of biosurfactant by fermentation of soybean oil or refinery wastes, such as acid oil, deodorizer distillate and soap stock. A three-layer neural network with 10 neurons in the input layer, 8 neurons in the hidden layer and 4 neurons in the output layer was developed for each substrate. The accuracy of the performance of the proposed neural network for the predicted dynamic model of biosurfactant production was confirmed with an  $R^2$  of 0.971, RMSE of 0.198 and SE of 0.041. The sensitivity analysis shows that the residual oil has the highest effect on the biosurfactant. In addition, the sensitivity results of the developed neural network model, as compared to those of the experimental data, are acceptable with a percent error of less than 4%.

### REFERENCES

- Gautam, K.K. and V.K. Tyagi, 2006. Microbial surfactants: A review. *J. Oleo. Sci.*, 55: 155-166.
- De Lima, C.J.B., E.J. Ribeiro, E.F.C. S ervulo, M.M. Resende and V.L. Cardoso, 2009. Biosurfactant Production by *Pseudomonas aeruginosa* grown in residual soybean oil. *Appl. Biochem. Biotechnol.*, 152: 156-168.
- Lotfabad, T.B., H. Abassi, R. Ahmadkhaniha, R. Roostaazad, F. Masoomi, H.S. Zahiri, G. Ahmadian, H. Vali and K.A. Noghabi, 2010. Structural characterization of a rhamnolipid-type biosurfactant produced by *Pseudomonas aeruginosa* MR01: Enhancement of di-rhamnolipid proportion using gamma irradiation. *Colloids Surf. B*, 81: 397-405.
- Omar, R., M.A. Abdullah, M.A. Hassan, M. Rosfarizan and M. Marziah, 2006. Kinetics and modeling of cell growth and substrate uptake in centella asiatica cell culture. *Biotechnol. Bioprocess Eng.*, 11: 223-229.

5. Sansonetti, S., T.J. Hobley, V. Calabrò, J. Villadsen and G. Sin, 2011. A biochemically structured model for ethanol fermentation by *Kluyveromyces marxianus*: A batch fermentation and kinetic study. *Bioresour. Technol.*, 102: 7513-7520.
6. Medina-Moreno, S.A., D. Jiménez-Islas, J.N. Gracida-Rodríguez, M. Gutiérrez-Rojas and I.J. Díaz-Ramírez, 2011. Modeling rhamnolipids production by *Pseudomonas aeruginosa* from immiscible carbon source in a batch system. *Int. J. Environ. Sci. Tech.*, 8(3): 471-482.
7. Assidjo, N.E., D. Akaki, B.K. Yao and Y.T. Eboi, 2009. A Hybrid Neural Network Approach for Batch Fermentation Simulation. *Aust. J. Basic and Appl. Sci.*, 3(4): 3930-3936.
8. Hur, W. and Y. Chung, 2006. An Artificial Neural Network for Biomass Estimation from Automatic pH Control Signal. *Biotechnol. Bioprocess Eng.*, 11: 351-356.
9. Kittisupakorn, P., P. Tangteerasunun and P. Thitiyasook, 2005. Dynamic Neural Network Modeling for Hydrochloric Acid Recovery Process. *Korean J. Chem. Eng.*, 226: 813-821.
10. Koprinkova, P., M. Petrova, T. Patarinska and M. Bliznakova, 1998. Neural Network Modelling of Fermentation Processes: Specific Kinetics Rate Models. *Cybernetics and Systems: An International Journal*, 29: 303-317.
11. Lotfabad, T.B., M. Shourian, R. Roostaazad, A.R. Najafabadi, M.R. Adelzadeh and K.A. Noghabi, 2009. An efficient biosurfactant-producing bacterium *Pseudomonas aeruginosa* MR01, isolated from oil excavation areas in south of Iran. *Colloids Surf B*, 69: 183-193.
12. Müller, M.M., B. Hörmann, M. Kugel, C. Syldatk and R. Hausmann, 2011. Evaluation of rhamnolipid production capacity of *Pseudomonas aeruginosa* PAO1 in comparison to the rhamnolipid over-producer strains DSM 7108 and DSM 2874. *Appl. Microbiol. Biotechnol.*, 89: 585-592.
13. Jagessar, R.C. and L. Sooknundun, 2011. Determination of nitrate anion in waste water from nine selected areas of coastal Guyana via a spectrophotometric method. *Int. J. Res. Rev. Appl. Sci.*, 7(2): 203-212.
14. Wei, Y.H., C.L. Chou and J.S.H. Chang, 2005. Rhamnolipid production by indigenous *Pseudomonas aeruginosa* J4 originating from petrochemical wastewater. *Biochem. Eng. J.*, 27: 146-154.
15. Krose, B. and P.V.D. Smagt, 1996. An introduction of neural networks. University of Amsterdam, Netherland.
16. Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*. Prentice Hall Co.