



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Software Sensor Development using Radial basis Function for Estimation of Erythropoietin Concentration

¹R. Rashid, ¹S.R. Radzali, ²B. Abdul Rahman and ¹S.M. Mohamed Esivan

¹Department of Bioprocess Engineering, Faculty of Chemical and Natural Resources Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

²Inno Biologics Sdn Bhd, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Abstract: Measurement of biological variables in a process is a key to efficient control and supervision of the bioprocess. In a process of protein production such as erythropoietin (EPO), it is crucial but difficult to measure EPO concentration using direct or on-line measurements. EPO concentration is usually measured through laboratory analysis where expensive costs of test kit, tedious and long time analysis are the biggest obstacles. Artificial neural network software sensor was developed to estimate EPO concentration based on other measured variables such as biomass, substrate or by-product in EPO production. Radial Basis Function was utilized to map nonlinear mapping between the input and output parameters. This study deals with effect of input numbers and spread constant on radial basis performance. It is found that different number of inputs and spread constant significantly affect the performance of the predictive model. The high values of coefficient of determination, R from regression analysis also proved that this model successfully mapping the nonlinear relationship between the input and output variables

Key words: Artificial intelligence, neural network, modeling erythropoietin production, radial basis function, orthogonal least square

INTRODUCTION

Erythropoietin or known as EPO is a hormone produced by specialized renal cells or by regenerating human hepatic cells in our body (Eckardt, 1996). These cells release EPO when the oxygen level is low in the kidney. EPO then stimulates the bone marrow to produce more red cells and thereby increase the oxygen-carrying capacity of the blood.

Normal levels of EPO are in the range 0 to 19 mU mL⁻¹ (milliunits per milliliter). Lower than normal levels of EPO are found in chronic renal failure while elevated levels of EPO can be seen in polycythemia, a disorder in which there is an excess of red blood cells. Nowadays, synthetic EPO has been produced through recombinant DNA technology in mammalian cell culture. Recombinant EPO is secreted from genetically engineered mammalian cells in a fermentation process then recovered and purified as EPO bulk in a purification process to achieve the desired product characteristics specified by the manufacturer. Besides of the very complicated manufacturing process, the final step in

EPO production also hampered due to the difficulty to measure the product itself.

The ability to provide faster and easier measurement of fermentation variables is important for monitoring and minimizing product quality variability. However, EPO concentration is usually measured offline through laboratory analysis where expensive costs of ELISA test kit, tedious and long time analysis via chromatography column are the biggest obstacles. Accurate prediction of EPO concentration can make it easy to formulate the EPO bulk into dosage form.

Since 1980s, Artificial Neural Network (ANN) based software sensor started to gain trust as a predicting tool where desired output can be predicted based on information of input from readily available on-line measurement (Golobic *et al.*, 2000; Bernard *et al.*, 2000; Cheruy, 1997). In the case of no available sensor, a software sensor may be designed using sampling and off-line laboratory analysis. It is a modeling approach to estimate difficult-to-measure variables from easy-to-measure variables (Jianxu and Huihe, 2002).

Radial Basis Function (RBF) network is one type of ANN which is a popular alternative to Multi Layer Perceptron (MLP) in modeling, controlling and optimization in bioprocesses. RBF captured interest because of their attractive properties such as localization, interpolation, cluster modeling and quasi-orthogonally.

There have been many successful applications of RBF network as a predicting tool. Moghadas and Choong (2008) study on prediction of double layer grids' maximum deflection using neural network resulted that RBF is better than backpropagation in term of training time and approximation errors. RBF network was proposed as soft sensor for automatic gantry crane system where the developed RBF model has estimated the unmeasured state well and was robust to parameter variations (Solihin *et al.*, 2006). The objective of this study is to develop a software sensor based on artificial neural network (ANN) for prediction of EPO production variables. This data-driven modeling approach is employed to predict EPO concentration based on other measured variables such as biomass, substrate (glucose and L-glutamine) or by product (ammonia).

Thus, RBF network is selected for this study because it has major advantages over MLP due to its ability to be trained using established linear regression techniques, allowing fast convergence to the solitary global minimum for a given set of fixed hidden nodes parameters (Dacosta *et al.*, 1997). Besides, RBF is may be used with advantage for modeling of a system with limited number of experimental data is available (Lanouetta *et al.*, 1999).

The developed RBF model is of a great importance due to its ability to predict variables under varying conditions. This study discussed the effect of spread constant and inputs numbers on RBF predictive performance in predicting EPO concentration.

MATERIALS AND METHODS

Model development

Structure and principle of RBF network: Radial basis function network consist of three layers which is an input layer, a single layer of processing perceptron and an output layer. The single layer of processing perceptron has an activation function called basis function. The most commonly used basis function is Gaussian basis function. Structure of RBF network in its most basic form is shown in Fig. 1.

The input layer is made up of source nodes whose number is equal to the dimension of the input vector *u*. The second layer is the hidden layer which is composed of nonlinear units that are connected directly to all of the nodes in the input layer. *x* is the desired output.

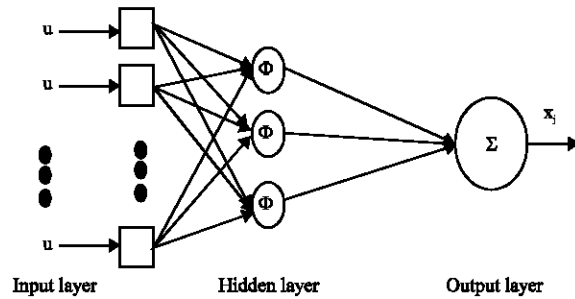


Fig. 1: Architecture of RBF network

Each hidden unit takes its input from all the nodes at the components at the input layer. The hidden units contain a basis function, which has the parameters center and width or spread. Φ is a basis function and *n* is no of cluster centers. The basis function is typically a Gaussian function, the spread σ corresponding to the variance which has a peak at zero distance and it decreases as the distance from the center increases.

In this work, Radial Basis Function Network model was developed using MATLAB 7.2, neural network toolbox. A feed forward radial basis function network with single hidden layer of nodes with Gaussian density function was chosen. MATLAB uses the Orthogonal Least Squares (OLS) algorithm to solve for the RBF centers and weights for the connections between the nodes in the hidden and output layers (Chen *et al.*, 1991). Other than specifying an error goal, the spread constant, σ , which determines the width of the receptive fields must also be specified respective to RBFN model development.

Data pre-processing: Data collection is obtained from Inno Biologics Sdn Bhd. All inputs and output data were preprocessed and normalized between zero and one using Eq. 1 to ensure each input variables provides an equal contribution in the network. This normalization method has been widely used in various studies (Vanek *et al.*, 2004; Choi and Park, 2001):

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_{norm} is the normalized value of variable *x*, x_{max} and x_{min} are variable maximum and minimum values, respectively.

RBF training and testing: There were four input and one output variables used in this work. The four input variables are biomass, glucose, L-glutamine and ammonia concentration, while the desired output for this model is erythropoietin concentration (Fig. 2).

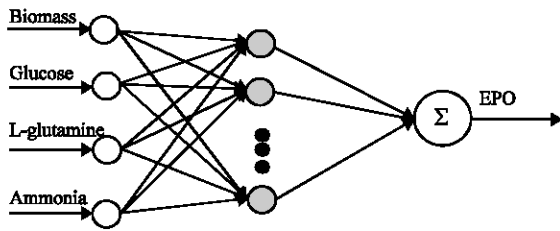


Fig. 2: Architecture of RBF network for predicting EPO concentration

Four models (A, B, C and D) with different number of input were presented in this study. Model A has one input (biomass), model B has two input variables (biomass and glucose), model C consists of three input variables (biomass, glucose and L-glutamine) and model D with four input variables (biomass, glucose, L-glutamine and ammonia).

Besides, the spread constant, σ was varied until the model obtained minimum error index. This step was conducted to investigate the effect of spread constant on RBF predictive performance. In this study, the predictive RBF model was designed using newrb function. This function can produce a network with zero error on training vectors. The function newrb takes matrices of input vectors P and target vectors T and a spread constant SPREAD for the radial basis layer and returns a network with weights and biases such that the outputs are exactly T when the inputs are P. Type of learning algorithm used in this newrb function is Orthogonal Least Square.

Model performance as expressed through Error index (EI): In the subsequent analysis, the RBF network performance is expressed throughout in term of error index, EI Eq. 2, because it provides a measure of suitable fitness of the model to the data (Rashid *et al.*, 2006):

$$EI = \sqrt{\frac{\sum(y - \hat{y})^2}{\sum y^2}} \quad (2)$$

y represents the experimental (real) value of output while \hat{y} is the predicted value.

Regression analysis between the network response and the corresponding target was performed to examine the network response in more details. Coefficient of determination, R was used as an indicator in this analysis.

RESULTS

Error Index (EI) and coefficient determination value (R) of training and testing set for each model were

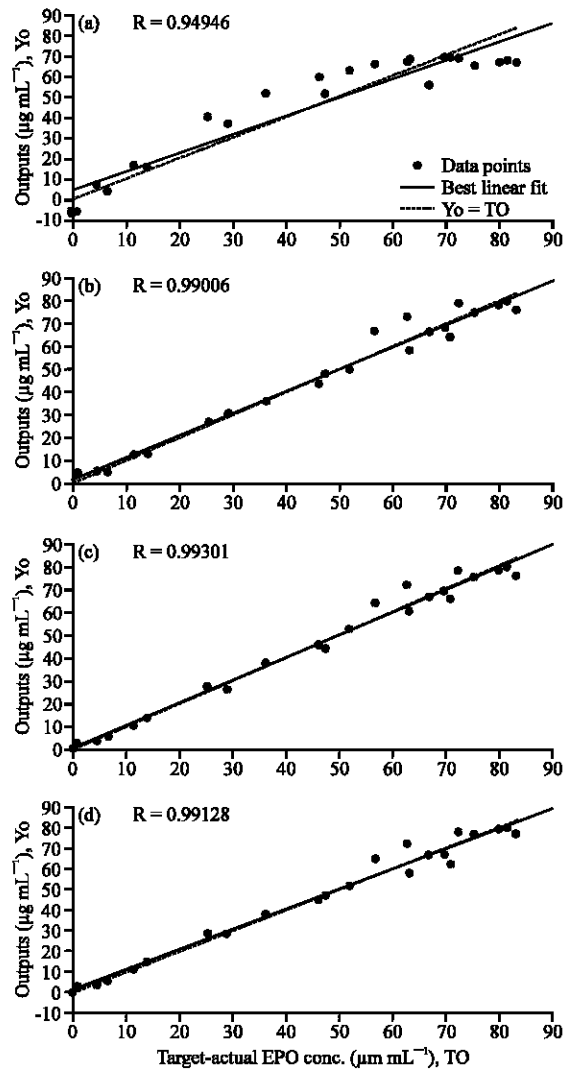


Fig. 3: (a, b): Regression analysis between predicted output and actual target for 1 input (model A) and 2 input (model B), respectively. (c, d) Regression analysis between predicted output and actual target for 3 input (model C) and 4 input (model D), respectively

collected during the simulation process and shown in Table 1 and 2. As shown, in Table 1 and 2, a network with spread constant, $\sigma = 4$ found to be an optimum RBF network because it fits for every model in term of EI percentage and R value.

A regression analysis of predicted output and actual target was performed to investigate the model precisely. Figure 3a-d illustrates the strong correlation between predicted value by the RBF model using the optimum spread constant and the actual value resulted from experimental data. The coefficient of determination, R for

Table 1: Error index and coefficient determination for different input numbers at selected spread constant, σ (Training set)

Selected σ	% EI (A)	R(A)	% EI (B)	R(B)	% EI (C)	R(C)	% EI (D)	R(D)
0.2	4.3	0.9657	7.95	0.8135	7.74	0.8242	10.11	0.6729
0.6	3.12	0.9736	3.22	0.9718	3.14	0.9732	3.13	0.9734
4	4.29	0.9495	1.92	0.9901	1.61	0.993	1.8	0.9913
10	4.32	0.9486	1.91	0.9902	1.59	0.9932	1.78	0.9915
22	4.33	0.9485	1.91	0.9902	1.6	0.9931	1.78	0.9915

Table 2: Error index and coefficient determination for different input numbers at selected spread constant, σ (Testing set)

Selected σ	% EI (A)	R(A)	% EI (B)	R(B)	% EI (C)	R(C)	% EI (D)	R(D)
0.2	0.59	0.9663	15.01	0.9385	22.33	0.7545	14.12	0.7566
0.6	2.01	0.9766	0.11	0.9125	11.32	0.883	11.8	0.8649
4	3.02	0.9674	7.57	0.8207	1.7	0.7977	0.28	0.7356
10	3.03	0.967	7.63	0.82	1.4	0.7948	1.38	0.7121
22	3.04	0.9669	7.64	0.8199	0.89	0.7753	1.56	0.7081

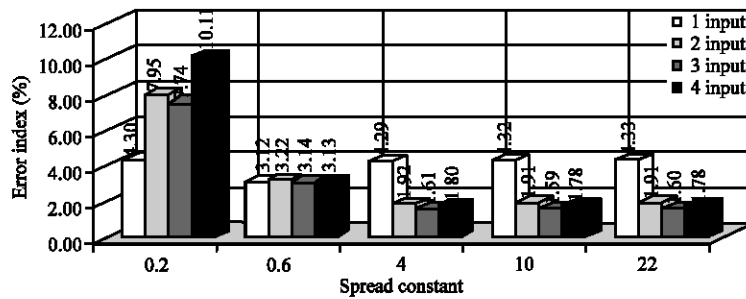


Fig. 4: Error index for different number of input at selected spread constant (training set)

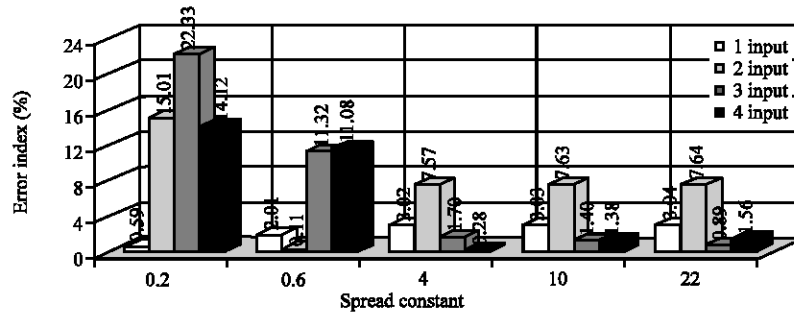


Fig. 5: Error index for different number of input at selected spread constant (testing set)

model A, B, C and D were 0.94946, 0.99006, 0.99301 and 0.99128 accordingly, which means that the model was successfully mapping the relationship between input and output variables.

DISCUSSION

Effect of spread constant, σ : Model A, B, C and D were trained with several spread constants (0.2, 0.6, 4, 10 and 22). As a result, it is found that smaller σ does not necessarily generate smaller EI and bigger σ also does not essentially resulted in higher error index. Figure 4 indicates that spread constant, $\sigma = 0.2$ produces higher error index for every model compare to spread constant of 0.6, 4, 10 or 22.

In Fig. 5, although $\sigma = 0.2$ generates very small error for model A ($E = 0.59\%$), but the other models resulted in very high EI (15.01, 22.33 and 14.12%). By setting the spread constant of testing set to 0.6, the lowest EI (0.11%) was obtained. However, only one and two input variable can generate small EI, while the bigger input variables gain slightly high number of EI. Thus, 0.2 and 0.6 are not fit to be selected as an optimum spread constant for this predictive model.

In this case, the spread constant of 0.2 and 0.6 may not large enough for the receptive fields to overlap one another to amply cover the whole input range. Nevertheless, it should not be too large that there is no distinction between the outputs of different nodes in the same area of the input space.

Effect of input number: As shown in Fig. 5, there was a case where the selected σ ($\sigma = 4$) gives very small EI (0.28%) at certain model (Model D testing set). However, when σ equal to 4 is applied to model A, B and C, high value of EI (1 to 7.57%) was obtained. This result shows that number of inputs affect the RBF predictive performance. At constant σ , small inputs number may gives high value of error index, which implies that error index value is proportional to the number of input variables used in the model.

CONCLUSION

In this study, RBF-based predictive model was proposed to predict EPO concentration based on measured variables such as biomass, glucose, L-glutamine and ammonia concentration. The best predictive model that successfully produced small error shows that *newrb* function that applied Orthogonal Least Square (OLS) training algorithm worked very well upon the centers and weights for the connections between the nodes in the hidden and output layers. Spread constant and number of inputs indeed affect the predictive performance where the optimum spread constant for this model is 4 and the best input number is three variables (biomass, glucose and L-glutamine concentration). Strong correlation between the input and output variables was indicated by high value of coefficient of determination, R where it proved that the model successfully mapping the nonlinear relationship between the input and output variables. Thus, offers a fast and reliable prediction of EPO concentration.

ACKNOWLEDGMENTS

We are thankful to Inno Biologics Sdn Bhd for their contribution on data collection and to Ministry of Science Technology and Innovation (MOSTI) for financial support.

REFERENCES

Bernard, O., Z. Hadj-Sadok and D. Dochain, 2000. Software sensor to monitor the dynamics of microbial communities: Application to anaerobic digestion. *Acta Biotheor.*, 48: 197-205.

Chen, S., C.F.N. Cowan and P.M. Grant, 1991. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Networks*, 2: 302-309.

Cheruy, A., 1997. Software sensors in bioprocessing engineering. *J. Biotechnol.*, 52: 193-199.

Choi, D.J. and H. Park, 2001. A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. *Wat. Res.*, 35: 3959-3967.

Dacosta, P., C.Kordich, D. Williams and J.B. Gomm, 1997. Estimation of inaccessible fermentation states with variables inoculum sizes. *Artifi. Intell. Eng.*, 11: 383-392.

Eckardt, K.U., 1996. Erythropoietin production in liver and kidneys. *Curr. Opin. Nephrol. Hypertens.*, 5: 28-34.

Golobic, I., H. Gjerkes, I. Bajsic and J. Malensek, 2000. Software sensor for biomass concentration monitoring during industrial fermentation. *Instrumentation Sci. Technol. Des. Appl. Chem. Biotechnol. Environ. Sci.*, 28: 323-334.

Jianxu, L. and S. Huihe, 2002. Soft sensing modeling using neurofuzzy system based on rough set theory. *Proceedings of the American Control Conference, (ACC'02), Alaska*, pp: 543-548.

Lanouetta, R., J. Thibault and J.L. Valade, 1999. Process modeling with neural network using small experimental datasets. *Comput. Chem. Eng.*, 23: 1167-1176.

Moghadas, R.K and K.K. Choong, 2008. Prediction of double layer grids' maximum deflection using neural networks. *Am. J. Applied Sci.*, 5: 1429-1432.

Rashid, R., H. Jamaluddin and N.A.S. Amin, 2006. Empirical and feed forward neural networks models of tapioca starch hydrolysis. *Applied Artifi. Intell.*, 20: 79-97.

Solihin, M., I. Wahyudi and A. Albagul, 2006. Development of soft sensor for sensorless automatic gantry crane using RBF neural networks. *Proceedings of 2nd IEEE International Conference on Cybernetics and Intelligent Systems, (ICCIS'06), Bangkok*, pp: 80-85.

Vanek, M., P. Hrnec, J. Vovsik and J. Nahlik, 2004. Online estimation of biomass concentration using a neural network and information about metabolic state. *Bioprocess. Biosyst. Eng.*, 27: 9-15.