Neural and Hybrid Modelling of Biotechnological Process

Elena Chorukova, Ivan Simeonov

The Stephan Angeloff Institute of Microbiology,

Bulgarian Academy of Sciences,

Acad. G. Bonchev St., Bl. 26, 1113 Sofia, Bulgaria,

E-mail: elena@microbio.bas.bg, issim@microbio.bas.bg

Abstract: The paper deals with neural and hybrid modelling of the three main types of biotechnological processes – batch and fed-batch processes for the enzyme superoxide dismutase (SOD) production and continuous process of anaerobic digestion (AD) of organic wastes. Neural models for batch SOD production with and without dissolved oxygen regulation have been developed. A hybrid model for fed-batch process for SOD production and neural model for continuous AD process have been developed as well. The most appropriate network architecture in all cases is with one hidden layer with hyperbolic tangent activation function and output layer with linear activation function.

Keywords: SOD production, anaerobic digestion, laboratory experiments, neural models, hybrid models

Elena Chorukova received her MSc degree in Bioengineering from the Technical University of Sofia, Bulgaria, in 1994. Since 2000 she has been a Research Associate at the Bulgarian Academy of Sciences, in the research group "Mathematical Modeling and Computer Science" of the Stephan Angeloff Institute of Microbiology. She is the author/co-author of about 30 scientific papers published in international journals and conference proceedings.

Ivan Simeonov received his MSc and PhD degrees in Control Engineering from the Technical University of Sofia, Bulgaria, in 1971 and 1981, respectively. Since 1986 he has been a Senior Research Associate with the Bulgarian Academy of Sciences, where he is presently the Head of the research group "Mathematical Modeling and Computer Science" at the Institute of Microbiology. He has been an Associate Professor at the Technical University of Sofia (1988-1999), Technical University of Gabrovo (2000-2005) and New Bulgarian University (since 2005). He is the author/co-author of over 160 scientific papers published in international journals and conference proceedings.

1. Introduction

The complexity and the non-linear dynamics of biotechnological processes require sophisticated methods for their modelling [1, 12]. Such methods should be able to update their knowledge and refine the model through interaction with the environment, that means the systems should have an ability of learning for a short time.

Theoretical results indicate clearly the advantages of Artificial Neural Networks (ANNs) when used especially for real data learning applications. Neural models are useful for many biotechnological applications [1, 2, 3].

On the basis of laboratory experiments and knowledge concerning the process for biosynthesis of the enzyme SOD, deterministic mathematical models have been developed [4]. However, it is very difficult to choose an appropriate non-linear structure for deterministic models including dissolved oxygen (DO). The ANN approach may be more useful in this case.

The anaerobic digestion (AD) has been long an object of mathematical modelling with deterministic approach [5], however parameter estimation of these non-linear models is a very hard problem [6]. That is why recently the instrument of ANNs has been used for AD modelling [7, 8, 11].

The objective of this study is to investigate the ability of the ANN approach for modelling of the three main types of biotechnological processes – batch, fed-batch and continuous. The studied processes have been as follows: batch and fed-batch processes for biosynthesis of the enzyme superoxide dismutase (SOD) and continuous process of AD of organic wastes.

2. Experimental Studies

Batch and fed-batch processes for the enzyme SOD production

Superoxide dismutases are metalloenzymes containing Mn, Fe or Cu/Zn in the active site, which play an important role in protecting aerobic organisms against oxidative damage [9]. Medical application of SODs is hindered by their limited availability.

The fungal strain *H. lutea 103* from the Mycological Collection of the Institute of Microbiology has been used in aerobic cultivation and maintained at $4^{o}C$ on beer agar, pH=6.3. Cultivation has been performed in a 3-L and 12-L bioreactors equipped with a pH-monitoring system, an automatic DO monitoring and controlling system. The compositions of the culture media has been as described earlier [9]. The cultures have been grown at $30^{o}C$ for 120 h.

Two types of batch processes including DO measurement has been performed: a) unregulated DO processes (only measurement of DO); b) regulated DO processes (measurement and regulation of DO at different levels). The fermentation parameters under unregulated DO conditions have been at an impeller speed of 500 rpm and an air flow of 1 vvm. In this case, the changes in DO during fermentation have been only measured. For regulated DO culture, the system aeration and impeller speed has been regulated to produce 20, 35, 50 or 60% oxygen saturation of the liquid.

For the fed-batch process glucose has been added at an interval of 3 h, starting in the 12^{th} hour from the inoculation, in concentrations of 0.75, 1.0, 1.5 and 2.0 gL^{-1}. The batch processing culture without glucose addition has been used as a basis.

Continuous process of AD

The anaerobic digestion of organic wastes is an interesting and perspective BTP with two positive aspects - ecological (depollution of wastes with high concentration) and energetical (biogas production).

Multiple experimental studies have been carried out on the laboratory-scale biogas plant of The Institute of Microbiology - The Bulgarian Academy of Sciences, with different organic wastes from farms and food industry [5]. Appropriate control actions (step- or pulse-wise) have been applied with acetate, glucose and different dilution rates for modelling and identification purposes.

Laboratory experiments have been carried out in 1-L and 2-L anaerobic bioreactors of continuously stirred-tank type. All reactors have been equipped with automatic control systems to maintain mesophilic conditions (temperature 34 ± 0.5 ºC) and shielded against light. To measure the volume of the obtained biogas, every bioreactor has been provided with water-replacement gas-holder. The pH in all bioreactors has been kept in the interval 6,5 ÷ 7,3. The analysis of chemical oxygen demand (COD), glucose concentration and concentration of the polluting organics have been made and used in modelling.

3. Neural Network Training and Simulation

For developing the neural models, the specialized software package NNSYSID20 for MATLAB 6.5 [10] have been used during the training, in which the following general regularized mean square error criterion is minimized:

$$W_N(\theta, Z^N) = \frac{1}{2N}\sum_{t=1}^{N}\left[y(t) - \hat{y}(t,\theta)\right]^2 + \frac{1}{2N}\theta^T B\theta \qquad (1)$$

where B is a positive definite matrix (most often $B=\alpha I$, $\alpha>0$); θ is a weight coefficients vector (matrix). The training algorithm has been based on the Levenberg-Marquardt method [10].

Comparative studies of some models of the type “black box” (including those using ANNs) show [11] that ARX (autoregressive with external input) models in most cases are superior to the other models. Neural model with ARX-structure shows the best results in our study as well, i.e. the model has feedback through the choice of regressors, which means that the networks become *recurrent*: future network inputs will depend on present and past network outputs. For this kind of dynamical models, it is known that the inputs to the network must be preceding data (at times (t-k), where k is determined experimentally) for the real physical inputs and for the measured outputs. An ARX-structure of first order for input and output data has been selected. Neural networks with different topologies have been studied. Topology of two layers, with activating functions - hyperbolic tangent for the intermediate layer and a linear one for the output layer has been selected. Automatic software pruning of the neural models has been performed only for those with one input and one output due to the limitations of the target software tool NNSYSID20.

The developed models validation has been performed in two ways: a) models testing with sets of data not included in the training process; b) analysis of some statistical characteristics, considering the autocorrelation functions (ACF) of model prediction errors and the cross-correlation coefficients curves (CCC) between the inputs and the prediction errors included in the target software tool NNSYSID20.

4. Results and Discussion

4.1. Modelling of batch process for the enzyme sod production

Models of the kinetics of growth and biosynthesis of the enzyme SOD for batch cultivation of *H. lutea 103* have been developed using deterministic and ANN approaches.

Models including the main process variables

The main process variables are: biomass concentration (X), limiting substrate (glucose) concentration (S) and enzyme SOD concentration (E). Data from four experiments for these variables in a 3-L bioreactor have been used for modelling – three for developing and one for validating the models. Experimental data have been very rough and that is why they have been filtered with Butterworth filter and interpolated with cubic spline.

Deterministic model

It is well known that the behaviour of the process is strongly non-linear. Therefore in deterministic models the coefficients estimation is very difficult to be resolved. The process has been described with the set of ordinary differential equations (2)-(4). All values of its coefficient are shown in Table 1 [4].

$$\frac{dX}{dt} = k_1SX - k_2X \tag{2}$$

$$\frac{dS}{dt} = -k_3X - k_4S^2X \tag{3}$$

$$\frac{dE}{dt} = k_5SX^2 - k_6X - k_7X^2 \tag{4}$$

Table 1

Coefficient	Value
k1	0.003692
k2	0.0008579
k3	0.001865
k4	0.0007156
k5	0.001
k6	0.00834
k7	0.001216

ANN model

An ANN model of this process has been developed. It possesses multi-input, multi-output (MIMO) structure with the main process variables: X, S and E as inputs. The outputs of the model are the same variables in the next moment of time. Study of the influence of the number of neurons has been made with number of neurons from 6 to 20. The best results have been obtained with 9 neurons. Study of the influence of the number of iterations (from 100 to 2500) for training of ANN has been made. The best training has been obtained with 1000 iterations. The influence of the number of the experimental data for training the ANN on the quality of the model has been studied. It shows that at least 22 input-output data pairs for training are necessary.

Some results concerning the performance of both models (neural and deterministic) are presented on **Fig. 1**. The relative mean square errors (RMSE) of both models calculated by (5)

$$RMSE=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i^{\exp}-Y_i^m}{Y_i^{\exp}}\right)^2 \tag{5}$$

where Y^{exp} designate desired outputs, Y^m designate model outputs are presented in Table 2. The RMSEs show that the neural model predicts better than the deterministic one and its design is much easier.

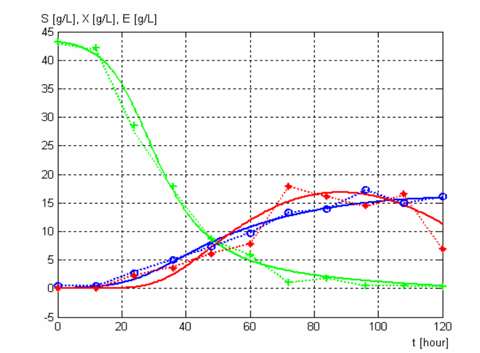

Figure 1. Testing results for deterministic and neural model for batch SOD production (real data for $^+$S, ^{o}X, *E; – deterministic model; neural model)

Neural models including unregulated DO

First of all we have developed neural models with single-input, single-output (SISO) structure, where DO is an input; X, protein concentration (P) and E are outputs respectively. These models have been evaluated from repeated experiments using seven subsequent runs (3 in 12-L bioreactor and 4 in 3-L bioreactor). Six experiments have been used for training and one for testing the networks. All data are shown on **Fig. 2.**

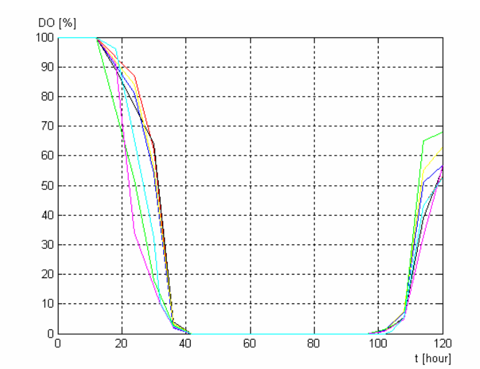

Figure 2. Inputs of neural models of batch process for SOD production including unregulated DO

The best results have been obtained with 6 neurons in the intermediate layer, a regression structure of the input layer including one previous input and two previous outputs and 500 training iterations. The models validity has been proved with new set of experimental data. In all studies, errors on testing have been commensurable with those on training, which is an important sign of adequacy of the obtained models. ACF and CCC for training and testing are practically within the confidence 95% intervals. Data for experimental and predicted outputs from testing the neural models are shown on **Fig. 3**. They are very close one to another and the RMSE values are presented in Table 2, e.g. the neural modeling is quite appropriate for this case.

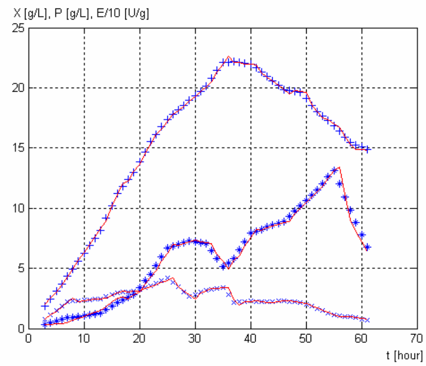

Figure 3. Testing results of neural models for batch SOD production including unregulated DO (real data for $^{+}$X, ^{x}P, *E/10; – neural models)

In order to study the ability of ANN approach to model more complicated relations of this process we have developed as well multi-dimensional neural model with MIMO structure, where the inputs are DO and S; the outputs are X and E. The validating results show that testing error is commensurable with this in the SISO case.

Neural models including regulated DO

As in the previous case, the input of the neural models in this section is DO, outputs are X, P and E respectively. The neural models have been evaluated from experiments using DO culture system aeration regulated to produce step inputs at 20, 35, 50 and 60% oxygen saturation of the bioreactor. For each of the 20, 35 and 50% oxygen saturation two experiments have been performed and one experiment for the 60% oxygen saturation. Six of these experiments have been used for training and one for testing the neural models. Best results have been obtained with 6 neurons in the intermediate layer, a regression structure of the input layer including two previous inputs and three previous outputs and 500 training iterations. In all studied cases, errors on testing have been commensurable with those on training. ACF and CCC for training and ACF for testing are within the 95% confidence intervals, but CCC for testing have been outside the intervals. Data for experimental and predicted outputs from testing the neural models are shown on **Fig. 4.** They are very close one to another and the RMSE values are presented in Table 2, e.g. the neural modelling is quite appropriate for this case.

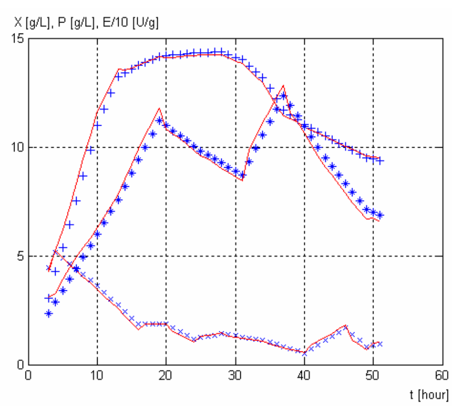

Figure 4. Testing results of neural models for batch SOD production including regulated DO (real data for $^{+}$X, ^{x}P, *E/10; – neural models)

4.2. Modelling of fed-batch process for the enzyme sod production

Data processing

Real experimental data of X, S and E are very rough as it is visible form **Fig. 5-7**. That is why a preliminary filtering has been applied using Butterworth filter. On Fig. 5-7 real and filtered experimental data for S, X and E (1.5 g/L and 2.0 g/L glucose feeding) are shown. **Fig. 8** shows filtered data for E with different feeding glucose concentrations. With the increasing of feeding glucose concentrations, the maximum of the enzyme activity increases very fast and move away on right in the time. This fact makes modelling of the fed-batch SOD production very difficult.

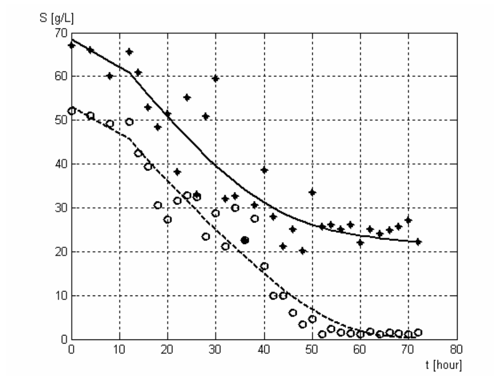

Figure 5. Real experimental data (^{o}1.5 g/L feed; *2.0 g/L feed) and filtered data (--1.5 g/L feed; – 2.0 g/L feed) for S (fed-batch SOD production)

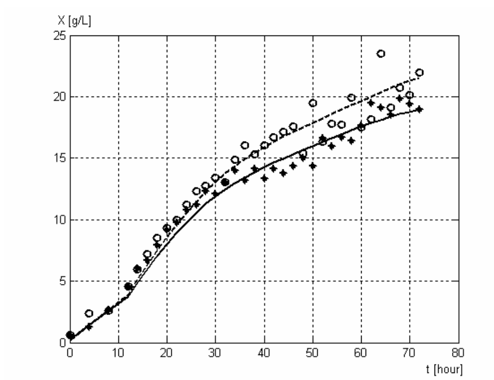

Figure 6. Real experimental data (^{o}1.5 g/L feed; *2.0 g/L feed) and filtered data (--1.5 g/L feed; – 2.0 g/L feed) for X (fed-batch SOD production)

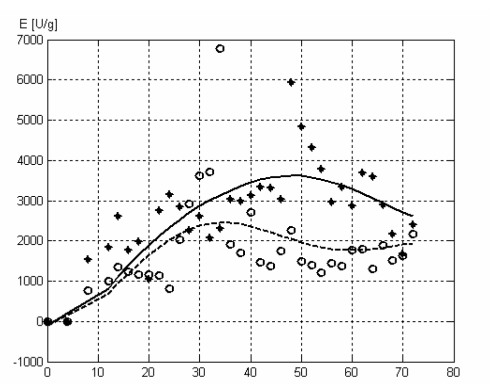

Figure 7. Real experimental data (^{o}1.5 g/L feed; *2.0 g/L feed) and filtered data (--1.5 g/L feed; – 2.0 g/L feed) for E (fed-batch SOD production)

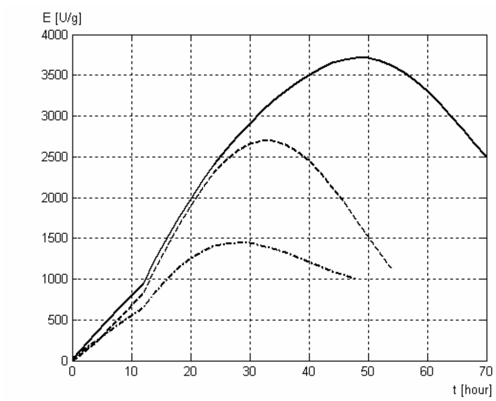

Figure 8. Filtered experimental data (-·-1.0 g/L feed; --1.5 g/L feed; – 2.0 g/L feed) for E (fed-batch SOD production)

Hybrid model

Developing models of the kinetics of growth and biosynthesis of the enzyme SOD for fed-batch cultivation with glucose as limiting substrate of *H. lutea 103* using deterministic and ANN approaches have not brought to good results. Therefore we have tried the following approach. We have used a general model for fed-batch biotechnological process (6)-(8) [12] with replacement the specific rates (of substrate consumption - ρ, biomass growth - μ and enzyme production - ξ) with neural networks.

$$\frac{dX}{dt} = \mu X \tag{6}$$

$$\frac{dS}{dt} = -\rho X + \frac{F}{V} S_{in} \tag{7}$$

$$\frac{dP}{dt} = \xi X - kP \tag{8}$$

At the first step we have tried to make hybrid model by replacing μ with ANN. At the second step μ and ρ have been replaced with ANNs. The best result has been obtained by replacing all specific rates (μ, ρ and ξ) by three different ANNs. These networks possess MISO (multi-input, multi-output) structures, whereas X and S are inputs of each of them and the specific rates are outputs. They are *recurrent* - feedbacks exist from outputs to inputs layer.

Filtered data for feeding with 1.0 g/L and 2.0 g/L glucose concentration have been used for modelling and these for feeding with 1.5 g/L glucose concentration – for testing. Testing results of S, X and E are shown on **Fig. 9.** The calculated values for RMSE in this case are presented in Table 2. They show that the hybrid model is appropriate for modelling the process.

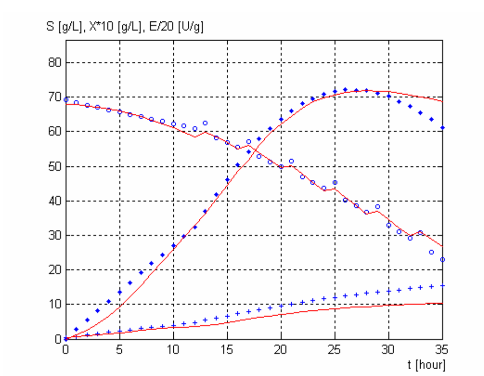

Figure 9. Testing results of hybrid model for fed-batch SOD production (filtered data for ^{o}S, $^{+}$X.10, *E/20; – hybrid model)

Table 2

Model output	S	X	P	E
RMSE of deterministic model of main variables of batch SOD production	0.7826	0.1038	-	0.1846
RMSE of neural model of main variables of batch SOD production	0.0007	0.0016	-	0.0010
RMSE of neural models including unregulated DO of batch SOD production	-	0.0002	0.0029	0.0155
RMSE of neural models including regulated DO of batch SOD production	-	0.0058	0.0089	0.0056
RMSE of hybrid model of fed-batch SOD production	0.0017	0.0714	-	0.0131

4.3. Continuous process of AD

ANN models with generated data

In order to study the ability of ANN approach to model the continuous process of AD of organic wastes first of all we have developed neural models trained with data generated from the deterministic model described in [5]. The following neural models have been developed:

SISO models:

- the control action "dilution rate for the inlet organics" in the bioreactor (D_1) has been assumed as the input, biogas production (Q) as the output;
- D_1 has been assumed as the input, COD as the output;
- the control action "concentration of the feed glucose" in the bioreactor (S_{10}) has been assumed as the input, Q as the output;
- S_{10} has been assumed as the input, COD as the output;
- the immeasurable "on-line" perturbation "concentration of the polluting organics" in the influent of the bioreactor (S_{0i}) has been assumed as the input, Q as the output.

MIMO models:

- two inputs (S_{10} and S_{0i}) and one output (Q);
- two inputs (S_{10} and S_{0i}) and two outputs (Q and COD).

All of these models possess *recurrent* structure. Different step-wise input actions have been used on training and testing the networks. Practically no statical error exists and an insignificant dynamical error has been observed during the training. During the testing with different values of the input action, the statical and dynamical errors have been near zero.

Validation with real experimental data for neural model with input S_{10} and output Q has been performed. Some results are shown on **Fig. 10** with RMSE value of 0.0494. Validation results using both statistical characteristics are within the confidence interval. Consequently, the neural models are appropriate for modelling the process.

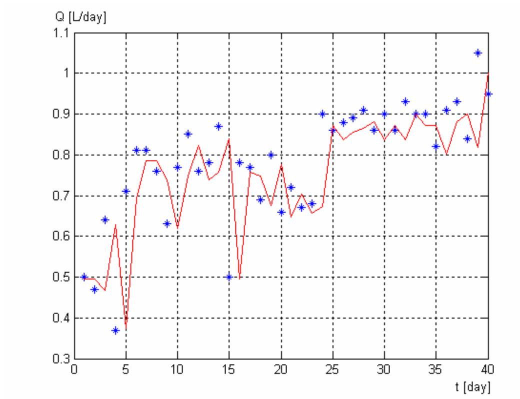

Figure 10. Testing results of neural model of AD with input S_{10} output Q, trained with generated data (* real data; – neural model)

ANN model with real experimental data

On the basis of the knowledge obtained from the theoretical study described above, investigations with real experimental data collected through the laboratory-scale biogas plant have been done as well. Neural model with input S_{10} and output Q with real changes of the input action as in the theoretical case have been developed. Testing with experimental data with step-wise and pulse-wise inputs have been performed. The testing results from step-wise input are shown on **Fig. 11** where circles designate the real data, and stars – the predictive network output. Experimental and model data are rather close to one another with RMSE value of 0.0008. ACF and CCC are within the admissible confidence intervals.

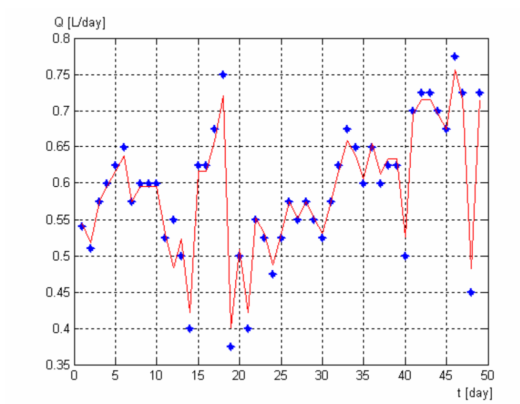

Figure 11. Testing results of neural model of AD with input S_{10} output Q, trained with real experimental data (* real data; – neural model)

5. Conclusions

Neural networks with different architectures have been studied during the modelling of all processes. The most appropriate network architecture in all cases is with one hidden layer with hyperbolic tangent activation function and output layer with linear activation function. Automatic software pruning did not improve the models.

Deterministic and neural models of the main variables (protein concentration, biomass concentration and enzyme activity) of batch process for the enzyme SOD production have been developed. The neural model predict better than the deterministic one.

Neural models for batch cultivation without DO regulation have been developed. They could be used for one-step-ahead prediction of the real time immeasurable state variables (protein concentration, biomass concentration and enzyme activity).

Neural models for batch cultivation with DO regulation using dissolved oxygen as a control input have been developed. They represent a good basis for developing of optimal control strategies of the process.

A hybrid model for fed-batch process for the enzyme SOD production has been developed, in which specific growth rates (for biomass growth, substrate consumption and enzyme formation) have been modelled using the ANN approach. They represent a good basis for developing of optimal control strategies of the process.

Neural models of the AD of organic wastes have been developed. They could be used as one-step-ahead predictor of the biogas flow rates for different dilution rates and for process control design.

REFERENCES

1. EIKENS, B., KARIM, M., SIMON, L., **Neural networks and first principle models for bioprocesses**, Proc. of the 14th IFAC congress, 1999, pp. 367-372.
2. NUCCI, E.R. et al., **Comparing the performance of multilayer perceptrons networks and neuro-fuzzy systems for on-line inference of *Bacillus megaterium* cellular concentrations**, Bioprocess Biosyst Eng., 30, 2007, pp. 429-438.
3. TE BRAAKE, H., **Neural Control of Biotechnological Processes**, PhD thesis, Technical University of Delft, 1997.
4. SIMEONOV, I., CHORUKOVA, E., DOLASHKA, P., ANGELOVA, M., **Modelling of the growth kinetics and the biosynthesis of the enzyme SOD in batch cultivation**, Automatics and Informatics 2000, Sofia, Oct. 24-26, 2000, pp. 77-79 (in Bulgarian).
5. SIMEONOV, I., GALABOVA, D., QUEINNEC, I., **Investigations and mathematical modelling of the anaerobic digestion of organic wastes with addition of electron acceptors**, 9th World Congress “Anaerobic Digestion”, Antverpen, Sept. 2-5, 2001, pp. 381-383.
6. SIMEONOV, I., **Mathematical modelling and parameters estimation of anaerobic fermentation processes**, Bioprocess Eng., 21, 1999, pp. 377-381.
7. STEYER, J.P. et al., **Neural Network modelling of a depollution process**, Bioprocess Engineering, No 6, vol. 23, 2000, pp. 727-730.
8. CAKMAKCI, M., **Adaptive neuro-fuzzy modelling of anaerobic digestion of primary sedimentation sludge**, Bioprocess Biosyst. Eng., 30, 2007, pp. 349-357.
9. ANGELOVA, M., DOLASHKA-ANGELOVA, P., IVANOVA, E., SERKEDJIEVA, J., SLOKOSKA, L., PASHOVA, S., TOSHKOVA, R., VASSILEV, S., SIMEONOV, I., HARTMAN, H.-J., STOEVA, S., WESER, U., VOELTER, W., **A novel glycosylated Cu/Zn-containing superoxide dismutase: production and potential therapeutic effect**, Microbiology UK, vol. 147, 2001, pp. 1641-1650.
10. NORGAARD, M., RAVN, O., POULSEN, N. K., **NNSYSID-Toolbox for system identification with Neural Networks**, Mathematical and Computer Modelling of Dynamical Systems, Vol. 8, No 1, 2002, pp. 1-20.
11. PREMIER, G.C. et al., **A comparison of the ability of black box and neural network models of ARX structure to represent a fluidized bed anaerobic digestion process**, Wat. Res., Vol.33. No 4, 1999, pp. 1027-1037.
12. VAN IMPE, J., et al., **Advanced Instrumentation, Data Interpretation and Control of Biotechnological processes**, Kluwer Acad. Publ., Dordrecht, 1998.