UDK 66.098
ISSN 0352–9193

professional paper

Mathematical Modeling of Bioprocesses by Neural Networks

Matematičko modeliranje biotehnoloških procesa s neuralnim mrežama

Ž. Kurtanjek

Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 41000 Zagreb, Croatia

Summary

Mathematical modeling of bioprocesses by artificial neural networks (ANN) is presented. Bioprocesses are considered as complex, nonlinear and dynamic multiple input / output systems (MIMO). Proposed is a general structure of ANN model composed of three serially connected subsystems: auto regression moving averages (ARMA), module for principal component analysis (PCA), and subsystem with layers of static neuron networks (NN) with feedforward pattern progression. The ARMA subsystem accounts for approximation of process dynamics by finite differences. The PCA module has two objectives: 1) rejection of measurement noise and 2) data compression by removing of collinearity between measured process patterns, i.e. reduction of a high dimension input vector to a few principal components. The NN provides highly adaptive interconnectivity between input and output patterns, and approximation of their nonlinear functional dependence. Parameters of neurons are adapted by use of conjugate gradient technique with the Ribiére-Polak-Powell algorithm for minimization of variance between the ANN model and measured output test patterns of a bioprocess. Applicability of ANN models in biotechnology is illustrated by models for prediction of protein secondary and tertiary structures based on amino acid sequences, process identification in production of penicillin, and the study of ANN internal model control (IMC) in industrial production of baker's yeast.

Sažetak

U radu je prikazan postupak matematičkog modeliranja biotehnoloških procesa s umjetnim neuralnim mrežama (ANN). Biotehnološki su procesi složeni, dinamički i nelinearni sustavi s više ulaznih i izlaznih veličina (MIMO). Predložena je opća struktura ANN sustava složenog od tri serijski povezana podsustava: sustav s pomičnim autoregresijskim srednjim vrijednostima (ARMA), sustav za određivanje glavnih komponenata (PCA) i sustav sa slojevima statičkih neurona (NN) s unaprijednim (»feedforward«) prijenosom informacija. ARMA podsustav omogućava aproksimaciju dinamike procesa s pomoću relacija s konačnim razlikama. Podsustav PCA ima dvojaku zadaću: 1) uklanjanje mjernih pogrešaka i kolinearnosti između mjerenih procesnih veličina i 2) kompresiju podataka, to jest sažimanje visoke dimenzije vektora ulaznih veličina na nekoliko glavnih komponenata. NN podsustav omogućava adaptivno povezivanje između ulaznih i izlaznih veličina i aproksimaciju njihove međusobne nelinearne funkcionalne zavisnosti. Tijekom učenja parametri neurona adaptiraju se primjenom postupka konjugiranih gradijenata i Polak-Ribiére-Powellova algoritma za minimaliziranje varijance između izlaznih veličina sustava ANN i procesa. Primjena ANN modela u biotehnologiji je ilustrirana modelima za predviđanje sekundarnih i tercijarnih struktura prema slijedu aminokiselina, identifikaciji procesa tijekom fermentacije penicilina i studijom adaptivnog upravljanja s ANN modelom u upravljačkom sustavu (IMC) u industrijskoj proizvodnji pekarskog kvasca.

Introduction

From a systems point of view, industrial processes based on microorganisms, in comparison with chemical processes, have several distinct properties: complex interactions, high dimension of state vectors, nonlinear dynamics, memory, and internal control mechanism of microorganisms which are adaptive to variable process conditions. The understanding of bioprocesses is limited by the availability of biosensors for on-line measurement of essential biological state variables such as concentrations of rate limiting intracellular species, morphological

Presented at the 2nd Croatian Congress of Food Technologists, Biotechnologists and Nutritionists, Zagreb, June 15–17, 1994

characteristics, and distribution of microorganism population. Mathematical models aimed to process control approximate systems of high dimensionality to models with few extracellular and intracellular variables which encompass input / output relations and enable effective control. Usually models for control purposes can be divided into two main classes: analytical and input / output models. Analytical models are derived from mass, energy and momentum balances with inclusion of approximate rate functions of intracellular reactions. Formally, they are given as a set of ordinary nonlinear differential equations with parameters estimated from laboratory and/or industrial experiments. Input / output models partially or completely disregard fundamental principles but provide reliable mapping of input patterns to output. Such models are formally given as auto moving averages (ARMA) equations, »fuzzy logic« relations, and artificial neural networks (ANN).

Artificial (computer) neural network models mimic biological neural systems which are able to memorize data and deduce rules for goal oriented behavior. The model building element is a neuron with numerous inputs and a single output with a nonlinear activation function. ANN system is composed of several layers of neurons through which patterns are propagated, from input to output layer. Information, i.e. data and rules, are built into a network through neuron parameters and neuron connectivity. Through a training process an ANN system is exposed to numerous prepared test examples of input and output patterns and its structure and parameters are changed until based on given criteria the system is optimally adapted to a real process. In view of the complexity of biological processes, modeling by ANN is very attractive especially for control engineering tasks in industrial production. The high potential of neural networks stems from their ability to integrate various forms of information commonly found in industrial practice, such as: on-line data from a computer automated measurement system, off-line data from manually conducted laboratory experiments, off-line microscope data, qualitative observations by production engineers, and knowledge of control rules by human experts. Formally, ANN can be exploited as on-line identification systems for the following purposes: 1) monitoring unmeasured biological parameters, 2) as single step and/or long range predictors, 3) can be included into on-line adaptive control algorithms, 4) may be applied for process optimization, and finally 5) can be utilized as knowledge expert systems assisting production engineers.

ANN structure and algorithms

A mathematical model of an ANN is based on a model of information processing at a single neuron. The most commonly used model *(1)* is given by the static sigmoid activation function:

$$y_j^i = \frac{1}{1+\exp\left[-(\vec{W}_j^i \cdot \vec{x}_i + \Theta_j^i)\right]} \qquad /1/$$

Neurons are collected in layers and layers are assembled in a network, so that a pattern entering a neuron on an input layer is spread through neurons at interior or hidden layers until it reaches a corresponding neuron at an output layer. In general terms, information consisting of data and mapping rules, is represented by the set of network parameters $\vec{\beta} = (\vec{W}^i, \Theta)$, $i = 1,2..n$ distributed in n layers. ANN performance is adapted in a sense of a postulated criterion to a set of selected and prepared test input and output patterns of a real system. In engineering applications it commonly applies the statistical criteria of minimal variance between a model and a real system defined by:

$$\sigma^2(\vec{\beta}) = \frac{1}{2} \cdot \sum_{k=1}^{N} (\vec{y}_k^i - \vec{t}_k)^T \cdot (\vec{y}_k^i - \vec{t}_k) \qquad /2/$$

The gradient of the function /2/ is determined analytically by the back propagation algorithm *(1)*. Here is proposed an original method for network training based on the conjugate gradient method with the Polak-Ribiére-Powell algorithm *(2)* given by:

$$\vec{\beta}_{l+1} = \min_{\lambda} \sigma^2 (\vec{\beta}_l + \lambda \cdot \vec{d}_l) \qquad /3/$$

At each iteration step /3/ one dimensional minimization is performed along the search direction:

$$\vec{d}_l = -\frac{\partial}{\partial \vec{\beta}_l} \sigma^2 (\vec{\beta}_l) + \gamma_l \cdot \vec{d}_{l-1} \qquad /4/$$

$$\gamma_l = \max \left\{ 0, \left(\frac{\partial}{\partial \vec{\beta}} \sigma^2(\vec{\beta}_l) \right)^T \cdot \frac{\left(\frac{\partial}{\partial \vec{\beta}} \sigma^2(\vec{\beta}_l) - \frac{\partial}{\partial \vec{\beta}} \sigma^2(\vec{\beta}_{l-1}) \right)}{\left\| \frac{\partial}{\partial \vec{\beta}} \sigma^2 (\vec{\beta}_{l-1}) \right\|^2} \right\} \qquad /5/$$

The method is utilized in a batch mode, i.e. with a complete set of N training patterns, and in practice it has been found to be very efficient for application on a standard PC computer *(3–5)*.

When ANN are used to model a time varying bioprocess, the input / output patterns have to account for approximation of process dynamics. In view of patterns obtained from on-line measurements of process variables, derivatives can be approximated by r-th order ARMA model giving input to ANN in the following form:

$$\vec{x}_k \leftarrow (\vec{x}_k, \vec{x}_{k-1} \cdots \vec{x}_{k-r}, \vec{y}_k, \vec{y}_{k-1}, \vec{y}_{k-r})^T \qquad /6/$$

where the right hand side of the relation /6/ contains real process measured input and output patterns which corresponds to an ANN input vector on the left.

Measured process data are always corrupted with pseudo-random noise and often with bias errors. Also process data are often linearly correlated. During the training procedure ANN can filter out random components, but biased errors and collinearity of patterns can degrade ANN prediction power. Data can be prefiltered and collinearity removed in a preprocessing stage by decomposition of input patterns to principal components (PCA). Matrix X of k input patterns can be decomposed

as a sum of matrixes defined as products of the score and loading vectors:

$$X = \sum_{j=1}^{n} \vec{s_j} \cdot \vec{p_j}^T \qquad /7/$$

The matrix of input patterns X must have each input variable normalized over a complete set of patterns, such that $E(\vec{x_k}) = \vec{0}, \sigma^2(\vec{x_k}) = \vec{1}$. The score $\vec{s_j}$ is j-th eigenvector of the covariance matrix $X \cdot X^T$, while the corresponding loading $\vec{p_j}$ is the eigenvector of $X^T \cdot X$. The matrix defined by the product of score and loading associated with the largest eigenvalue accounts for the maximum data covariance, the next pair for the maximum of the difference, and so forth. The number of required components to account for 95 % of the covariance is usually small compared to the total number of components of the input vector, thus leading to significant reduction of input variables and pattern compression *(5)*. Besides the importance of dimension reduction, there is significant gain due to noise rejection by principal component analysis (PCA). Training of ANN on principal components is more efficient as noise is already eliminated, and also results in increased predictive power since ANN does not attempt to adapt to measurement and process noise. The PCA can not eliminate bias and gross errors and they need to be detected and rejected by a separate algorithm, such as data reconciliation based on mass and energy balances applied in pseudo steady state models.

The complete structure of an ANN model is presented in Fig.1. It is composed of three subsystems, or modules, connected in a series. The first subsytem is an ARMA module dedicated for discretization of model differential equations into finite difference from time series of input and output patterns. The second module is PCA and it has the task of reducing the number of input variables to few principle components and at the same time to reject measurement noise present in patterns. The third subsystem is a static artificial neural network (ANN) which is able to learn interaction, or mapping, between input and output patterns. This is the essential subsystem as it has to deduce model functions from data available in a training set.

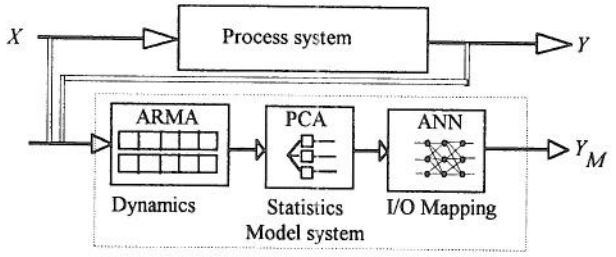

Fig. 1. Structure of the model of dynamic process composed of subsystems for autoregression moving averages (ARMA), principal component analysis (PCA), and artificial neural network (ANN)

Slika 1. Prikaz strukture modela dinamičkog procesa sastavljenog od podsustava s pomičnim autoregresijskim srednjim vrijednostima (ARMA), analizom glavnih komponenata (PCA), i umjetnom neuralnom mrežom (ANN)

ANN models in biotechnology

There are numerous examples in biology and biotechnology of computer modeling based on neural networks. An example is the software PROBE (PROtein prediction at BErkeley) for prediction of specific secondary and tertiary protein structural features from amino acid sequences *(6)*. The software includes several neural network modules for prediction of:

- percent helix, strand and coil in protein,
- presence of disulfide bonds,
- β–turns,
- folding patterns,
- surface exposure of protein residues.

The method is tested on known structure features of 273 proteins extracted from a data base of known crystal structures. The accuracy of the method is about 95 % for secondary and about 80 % for tertiary structural features. The NN models give improved accuracy by a factor 2 compared to models based on multiple regression technique.

Application of NN models in bioreactor monitoring has been tested in laboratory and industrial experiments for fed-batch and continuous operations. Results of comparative computer simulation studies of on-line prediction by NN models versus extended Kalman filter (EKF) method gave several important advantageous properties *(7)*. The accuracy of prediction of unmeasured biological properties by NN was better than for Kalman technique, the method is stable while EKF has a finite range of stability, and NN modules present negligible computer load while integration of filter equations in EKF can consume much computer time. The results have been proved in a number of cases in practice, such as in experiments with monitoring laboratory fermentations of *Bacillus thuringiensis (8)* and *Penicillium chrysogenum (9)* in industrial production.

In the laboratory experiments the input patterns were: type of inoculum, temperature, pH, accumulated process time, and optical density, while the output pattern was optical density at next sampling interval *(8)*. A sampling time of 1 hour was applied. Fermentations were inoculated with spores and vegetative cells and this is an important input pattern since it is related to duration of a process lag phase. A feedforward NN structure with static neurons was applied. The models with output optical density as a continuous variable and as a class with 9 qualitative grades were tested. The trained networks were used in a supervisory control structure for optimization of biomass concentration at next sampling interval. At each interval the input patterns were scanned for optimal optical density, and the values were used as input set points for a slave computer. Improvement of 10 % in productivity over classical PID control was obtained.

A study of neural network modeling of industrial penicillin production indicates that prospects for increase of productivity are significant *(9)*. Applied network topology consisted of four inputs, two hidden layers with six nodes at each, and a single output. The input patterns were two feed rates of nutrients, carbon evolution rate (CER), and accumulated process time. The output was biomass concentration. For optimization of

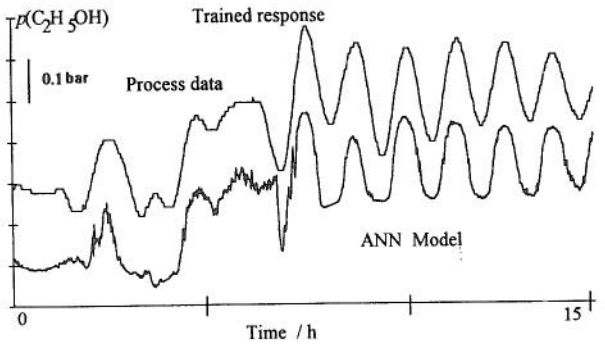

Fig. 2a. Comparison of the measured partial pressure of ethanol $p(C_2H_5OH)$ and the response by the neural network model (ANN) obtained in the network training procedure

Slika. 2a. Usporedba mjerenog parcijalnog tlaka etanola $p(C_2H_5OH)$ i odziva modela s neuralnom mrežom (ANN) dobivenog tijekom postupka učenja mreže

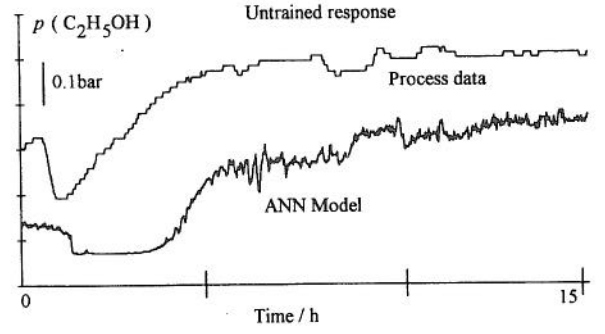

Fig. 2b. Comparison of the measured partial pressure of ethanol $p(C_2H_5OH)$ and the response by the neural network model (ANN) obtained in an untrained experiment

Slika. 2b. Usporedba mjerenog parcijalnog tlaka etanola $p(C_2H_5OH)$ i odziva modela s neuralnom mrežom (ANN) dobivenog u pokusu bez učenja

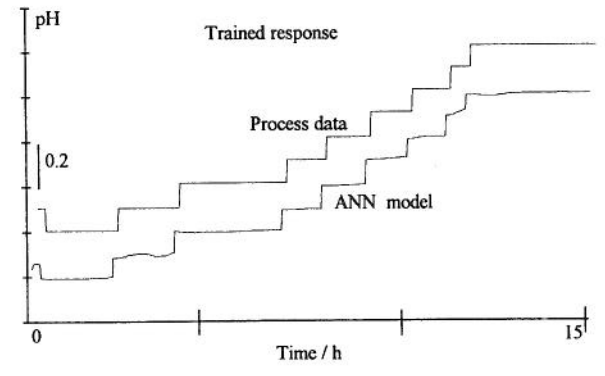

Fig. 3a. Comparison of the measured pH and the response by the neural network model (ANN) obtained in the network training procedure

Slika. 3a. Usporedba mjerenog pH i odziva modela s neuralnom mrežom (ANN) dobivenog tijekom postupka učenja mreže

penicillin production a series of two neural networks was considered. They are separately trained, the first for biomass and the second for penicillin concentration. Output from the first network is fed to the second. The model was tested on numerous industrial data logs and it was proven that the model can capture a nonlinearity of the process at the biomass growth phase, and also during the product synthesis phase. The developed model can be incorporated in a supervisory software and/or control for process optimization.

Control of bioreactor by ANN

Modeling by neural networks for control of industrial bioreactors has an outstanding perspective. It may open up the main bottle necks in control which are due to lack of specific and reliable biosensors and also due to only partly understood dynamics intracellular mechanisms of the microorganism. From a control perspective the main advantages of neural networks are their ability to integrate various forms of information such as on-line measurements and off-line laboratory data, their predictive power for identification of unmeasured state variables and parameters, and the ability to deduce rules from training data. By incorporation of NN modules in a control structure it becomes adaptive and nonlinear, and can be realized as: 1) identification and/or prediction of biological variables with information fed to classical controller, or 2) replacement of classical controller by inverse trained models *(10–12)*.

The potential of NN models for process control is demonstrated by the study of industrial baker's yeast fermentation. The process is automated and on-line measurements are available for: feed rates of nutrients, dissolved oxygen concentration, pH, ethanol partial pressure, rotation frequencies of electric motors for broth reflux, broth level, and temperature. Each on-line variable is sampled with a frequency of one reading per minute, while biomass concentration is determined off-line at intervals of one hour. Input patterns are also information on type of molasses and mass of inoculum. The process is controlled with classical PID loops for partial pressure of ethanol, pH, and temperature. Neural networks were trained on process data with active PID control. The objective of the study is to replace classical PID with adaptive control based on neural network models and to optimize process productivity. The NN system is composed of two independent modules dedicated for control of ethanol partial pressure and pH. The modules are trained for prediction of the state variables, and also as inverse models for prediction of the process manipulated variables. Each NN module has a topology given in Fig. 1. In each module only one hidden layer is utilized with up to seven internal nodes. The first order ARMA subsystem was applied. In Fig. 2. are shown the responses of the NN module trained for one step prediction of ethanol partial pressure. The responses in the figures are shifted by a constant amount in order to avoid curve overlaping and show fine qualitative agreements in details between measured and model information. For reasons of confideality values on ordinate scales are always given as relative quantities. The results in Fig. 2a. are obtained with the training set of patterns.

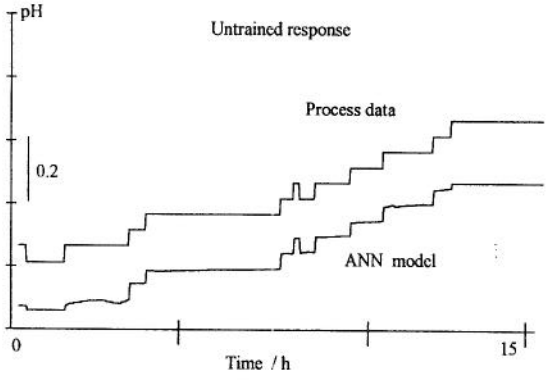

Fig. 3b. Comparison of the measured pH and the response by the neural network model (ANN) obtained in an untrained experiment

Slika. 3b. Usporedba mjerenog pH i odziva modela s neuralnom mrežom (ANN) dobivenog u pokusu bez učenja

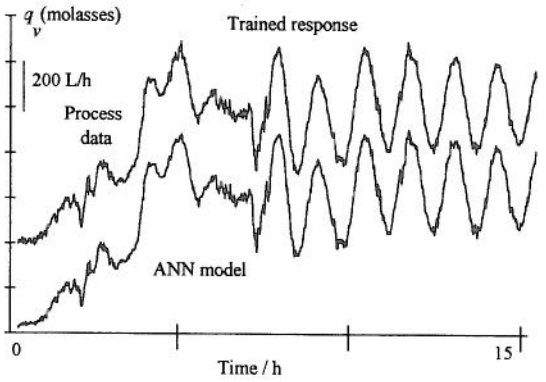

Fig. 4a. Comparison of the measured molasses flow rate q_v and the response by the neural network (ANN) inverse model obtained during training procedure

Slika 4a. Usporedba mjerenog protoka molase q_v i odziva inverznog modela neuralne mreže (ANN) dobivenog tijekom postupka učenja mreže

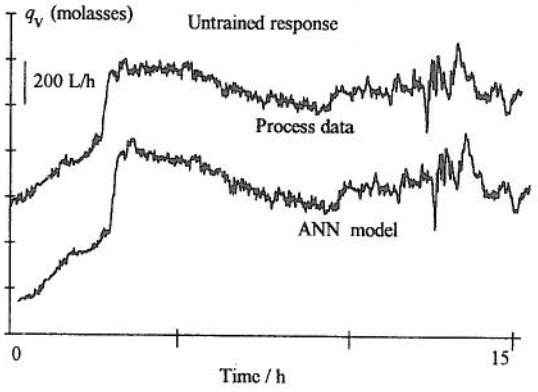

Fig. 4b. Comparison of the measured molasses flow rate q_v and the response by the neural network (ANN) inverse model obtained in an untrained experiment

Slika 4b. Usporedba mjerenog protoka molase q_v i odziva inverznog modela neuralne mreže (ANN) dobivenog u pokusu bez učenja

The oscillatory fermentation profile was obtained with unmatched parameters of the PID control loop, and it was selected for training as it provides more information (i.e. process variable variations) than is contained in a relatively stable fermentation. A response obtained in a pseudorandom sequence of input impulse perturbations would represent an ideal training set, but it is difficult to realize in an industrial process. In Fig. 2b. is given the response of the NN module trained at oscillatory conditions but with input patterns measured in a stable fermentation. In Fig. 3a. and 3b. are shown responses obtained with the NN module for prediction of pH. Besides the NN modules for one step prediction of ethanol pressure and pH, inverse modules were developed with the aim of including them directly in control loops. In these cases the output patterns are the corresponding manipulative variables in the individual control loops. Feed rate of molasses q_v*(molasses)* is the output for the module dedicated to control partial pressure of ethanol, and feed rate of ammonia $q_v(NH_3)$ is the output of NN for pH control. In Fig. 4. are shown the results obtained with the trained (Fig. 4a.) and untrained (Fig 4b.) set of data.

The results are evaluated by simple statistical analysis. The average relative error between the NN models and the measured data is less than 1 % for training experiments, and is lower than 3 % for the untrained experiments. The achieved accuracy in modeling is within the range of instrumentation error commonly found in industry.

Fig. 5 proposes a scheme for the internal model control (IMC) structure. The control is based on three independent feedback loops, a mass balance module, and a data base of optimal fermentations. In the control loops for ethanol pressure and pH classical controllers are replaced with the inverse NN modules while temperature is controlled with a PID. Outputs from the controllers are processed in the mass balance model for calculation of required feed rates for balanced yeast growth. The course of a fermentation follows a »model reference« trajectory in state space provided from a data base of optimal fermentations. Optimal fermentations can be determined analytically from mathematical optimization, or can be selected according to a preselected assumed cri-

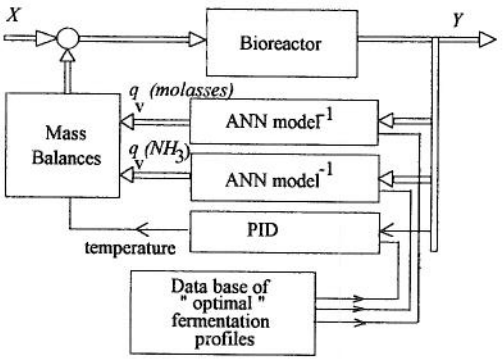

Fig. 5. Control structure with internal models (IMC) based on inverse neural modules for partial pressure of ethanol and pH, and PID controller for temperature

Slika 5. Upravljačka struktura s uključenim inverznim modelima (IMC) neuralnih mreža za parcijalni tlak etanola i pH, te PID regulatorom za temperaturu

terion from past data logs. The data base of optimal trajectories can account for different types of molasses, mass of inoculum, and other process data such as conditions of previous fermentations.

Conclusions

Applications based on neural networks in bioprocess engineering must account for nonlinear process dynamics. An ANN structure with ARMA approximations of input patterns is proposed to enable modeling of dynamic effects by layers of static neurons. Input patterns can be effectively preprocessed by principal component analysis (PCA) by which is accomplished a significant reduction of number of the inputs, pattern compression, and measurement noise reduction.

NN models can be effectively trained for prediction, such as structural features of proteins related to their biological activity from amino acid sequences, or in engineering tasks for prediction of the course of a fermentation and of on-line unmeasured biological variables.

In control of industrial bioprocesses NN models have outstanding prospects. They can circumvent major obstacles caused by lack of biosensors and uncertain analytical models. Control based on neural networks can be realized as a cascade structure with NN modules providing input information to slave PID controllers, or NN modules can be trained as inverse modules and directly incorporated in control loops.

In the study of modeling of industrial bioreactor for yeast production direct and inverse NN models are derived for ethanol pressure and pH with accuracy of average relative error of 1 % on trained data and 3 % with untrained experiments.

Acknowledgments:

This work was partially financially supported by the Ministry of Science and Technology of the Republic of Croatia, project 04–07–017.

Financial support from company »PODRAVKA« and assistance by Miroslav Repić, B. Sc., is gratefully acknowledged.

List of symbols
Popis oznaka

$\vec{d}$	direction vector vektor smjera
E	expected value matematičko očekivanje
n	total number of layers ukupan broj slojeva
N	total number of patterns ukupan broj uzoraka
p	partial pressure parcijalni tlak
$\vec{p}$	loading vector vektor težine
q_v	volumetric flow rate volumni protok
r	order of ARMA approximation stupanj ARMA aproksimacije
$\vec{s}$	score vektor vektor pogodaka
$\vec{t}$	training patterns uzorci za učenje
$\vec{x}$	vector of input patterns vekor ulaznih uzoraka
X	matrix of input patterns matrica ulaznih uzoraka
Y	matrix of output patterns matrica izlaznih uzoraka
y	output pattern from a single neuron izlazni uzorak iz pojedinog neurona
$\vec{W}$	vector of input gain coefficients vektor ulaznih koeficijenata pojačanja
$\vec{\beta}$	vector of all network parameters vektor cjelokupnih parametara mreže
λ	optimization parameter parametar optimiranja
γ	gain factor faktor pojačanja
σ^2	criteria function (variance) funkcija kriterija (varijanca)
$\vec{\Theta}$	vector of sensitivity threshold parameter vektor parametara praga osjetljivosti

Subscripts
Donji indeks

j	index of a neuron indeks neurona
k	index of patterns indeks uzoraka
l	iteration index indeks iteracija
M	model model

Superscript
Gornji indeks

o	output layer izlazni sloj
i	layer index indeks sloja
T	transposition operation operacija transpozicije

Literature

1. D. E. Rumelhart, J. L. McClelland: *Parallel Distributed Processing*, The MIT Press, Cambridge MA (1989) pp. 318–363.
2. M. J. D. Powell, *SIAM Review, 28* (1988) 487–500.
3. Ž. Kurtanjek, *Automatika, 33* (1992) 147–150.
4. Ž. Kurtanjek, *Comput. Chem. Eng. 18* (1994) 627–631.
5. Ž. Kurtanjek, Proceedings of 16th International Conference »Information Technology Interfaces«, Ed. V. Čerić and V. Hljuz Dobrić,14–17, June, Pula, Croatia, p. 277–281, 1994.

6. S. R. Holbrook, I. Dubchak, S.H. Kim, *BioTechniques, 14* (1993) 984–989.
7. J. Thibault, V. Van Breusegem, A. Cheruy, *Biotechnol. Bioeng. 36* (1990) 1041–1048.
8. Q. Zhang, J.F. Reid, J.B. Litchfield, J. Ren, S.W. Chang, *Biotechnol. Bioeng. 43 (1994) 483–489.*
9. C. Di Massimo, G. A. Montague, M. J. Willis, M. T. Tham, A. J. Morris, *Comput. Chem. Eng. 16* (1992) 283–291.
10. N. Bhat, T.J. McAvoy, *Comput. Chem. Eng, 14* (1990) 573– 583.
11. B.E. Ydstie, *Comput. Chem. Eng, 14* (1990) 583–599.
12. D.C. Psichogios, L.H. Ungar, *AIChE Journal, 38* (1992) 1499––1511.