Asian Journal of Applied Sciences (ISSN: 2321 – 0893)
Volume 03 – Issue 05, October 2015

Analysis of Sensor Data for Hydrogen Production in a Biofilm Photoreactor using Multilayer Perceptron Network

Slawomir Procelewski[1], Lucia Diez[2], Joanna Procelewska[3], Jangik(John) Park[4], Jeongyun (Lewis) Moon[5],Antonio Delgado[6*]

[1]Lehrstuhl fürStrömungsmechanik
Cauerstraße 4, 91058 Erlangen, Germany

[2]Lehrstuhl fürStrömungsmechanik
Cauerstraße 4, 91058 Erlangen, Germany

[3]Lehrstuhl fürStrömungsmechanik
Cauerstraße 4, 91058 Erlangen, Germany

[4]Director,Engineering Div. OCEANUS CO., LTD.
1480-7, Jung-dong, Haeundae-gu, Busan, Korea

[5]Manager, Engineering Div, OCEANUS CO., LTD.
1480-7, Jung-dong, Haeundae-gu, Busan, Korea

[6]Lehrstuhl fürStrömungsmechanik
Cauerstraße 4, 91058 Erlangen, Germany

[*]*Corresponding author's email: antonio.delgado [AT] fau.de*

ABSTRACT--- ***NeuralNetworks are one of the most appreciate techniques in the field of the analysis of the data set. In this paper the usage of multilayer perceptron networks (MLP) for the prediction of the hydrogen production from the sensor data is presented. The results with R^2 value of over 0.95 show clearly, that it is possible to build an effective system for the prediction of the hydrogen production and concentration rates based only on the data covering biofilm thickness.***

Keywords – Neural Networks, Hydrogen Production, Data Analysis

1. INTRODUCTION

The analysis of data set has been widely carried out in recent years by means of artificial intelligence [1], the bioscience field being one of the most successfully[2]. One of the most appreciate techniques is the implementation of Neural Networks (NN) thank to their capability of synthesizing the behavior of non-linear systems[3, 4]. Under this, the multilayer perceptron architecture (MLP) as universal approximator[5-11] has been widely studied.

In this paper, the usage of the MLP networksfor the prediction of hydrogen production in a biofilm photobioreactor (BPBR) based on provided experimental input data will be performed. Both input and output data are represented as the functions and the measurements are arranged in order that the process modeling is able to be approximated byNNs. Such reduction of information and the understanding of the correlations between the data gives an invaluable insight for the understanding of the process. The format of the data to a function implies that the relationships between the involved parameters is clearly exposed. Firstly, when considering a function, the relationships between all the parameters involved are clear. Moreover, the strategy of considering certain solutions would give an opportunity to exclude presence of other processes that may occurred during gathering of the data and have no influence on the results but generate the noise making the whole analysis more complex.However, in some situations, no clear correlation between input and output data can be found. In such cases, the artificial intelligence methods can provide the solutions of acceptable quality but significantly reduced computation time.

2. MATERIALS AND METHODS

1.1 Neural Networks

The fundamental unit of NNs is a neuron composed of one or more inputs, a main body consisting of weights, bias, weighting function and activation function and a single output leaving the neuron (**Error! Reference source not found.**, left). Two or more neurons can be combined to form a layer using a weighting matrix. Two layers are connected when the output of neurons of one layer becomes an input of neurons of another layer. The layer where the input values are provided is called the input layer and the layer generating the final output the output layer while the internal layers are neurons denoted as hidden layers. Depending on the network structure, the architecture and the learning methods, several categories of NNs can be defined.

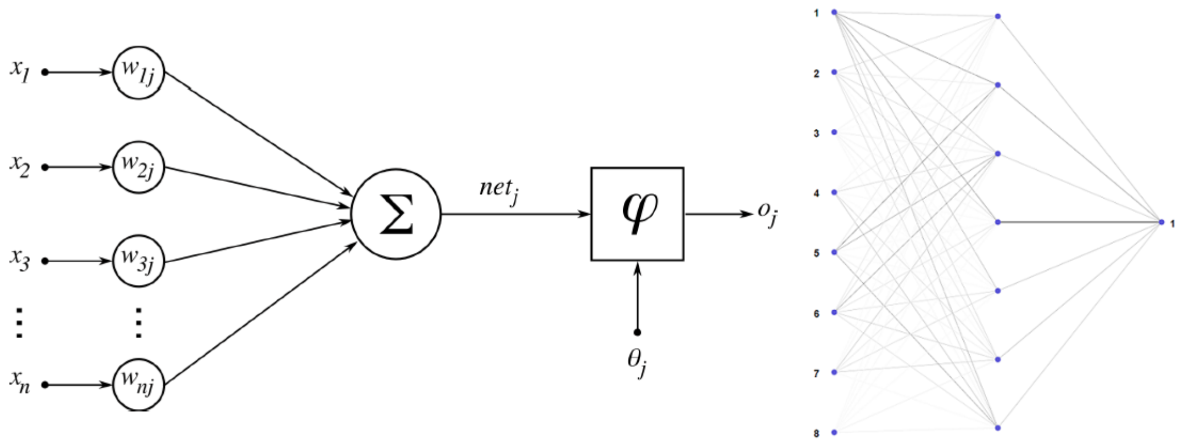

Figure1: Schema of the single neuron (left) and the neural network (right). The color and thickness of the lines connecting neurons represent the weight values.

A multilayer perceptron is a feedforward network, which means that connections between neurons do not form any cycles and the output of each neuron in the layer follows to the next layer, omitted layer being not allowed. Typically MLP consists of an input layer containing one neuron for each value of input vector, one hidden layer and an output layer containing one neuron for each value of the output vector. However, it is possible to define more than one hidden layer.

It is well known that a two-layered NNs, i.e. one thatdoes not have any hidden layers, is not capable ofapproximating generic nonlinear continuous functions[12]. On the other hand, four or more layerNNs are rarely used in practice because any Boolean function can be realized by a MLP with one hidden layer[6],fulfilling the requirements for the most applications. Therefore the known results in literature[5, 8, 11] are mainly built upon thethree-layered neural networks with one linear output node. The present work adopts this standard setting as well.

1.2 Data Preparation and Analysis

The monitoring of biofilm biohydrogen production process is very important to understand the effects of the biological and physical factors on biofilm metabolic activity and to optimize hydrogen production performance[13]. The biological information acquisition and subsequent analysis are very important to design the control system of a biofilm, because the biofilm growth, biofilm activity and hydrogen production performance are directly related to the biomass concentration, thickness and structure[14, 15].

In order to improve biofilm activity and obtain high hydrogen production performance of biofilm photobioreactors, Liao et al. proposed recently an optimization system for biofilm growth on the fiber-optic surface[16]. The proposed measurement system consisted of several types of sensors, among others biofilm thickness sensors. For this system, together with monitoring of the biofilm growth, the values of hydrogen concentration and production wereregistered. The measurements took place in two different setups: with and without biofilm thickness control apparatus. For the last case, a power controller acted to promote an additional shear stress in the biofilms. The biofilm thickness was investigated with five sensors, the sampling time interval being 36 minutes. The noise visible on the right chart of **Figure**2was caused by the strong exceeding of the sensor measurement range.However,no data point is discarded in the present work, in order to check whether this noise impacts the network behavior.

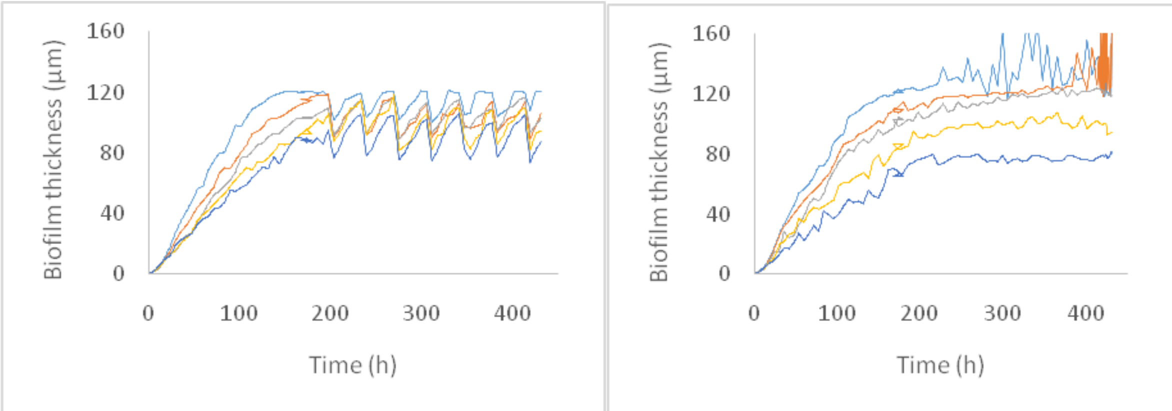

Figure2: Sensor data of biofilm thickness and distribution for controlled (left) and uncontrolled (right) setup from Liao et al.

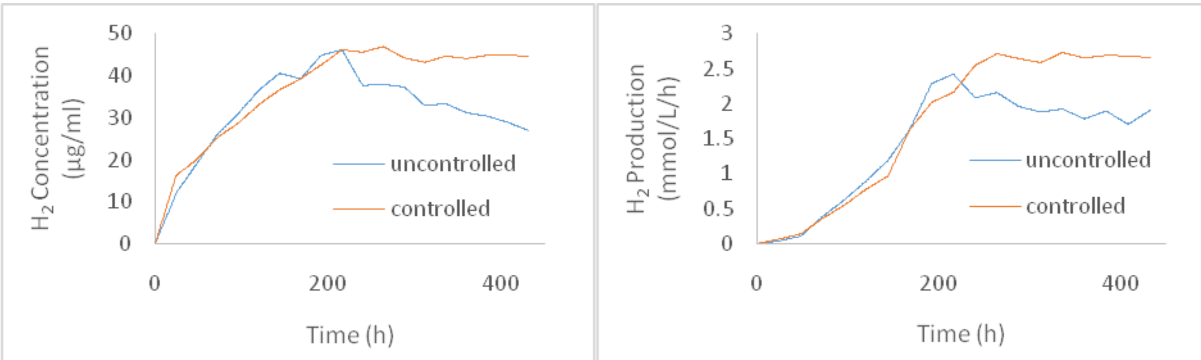

Figure3: Hydrogen concentration and production rates for the investigated systemfrom Liao et al.

To investigate the effects of the biofilm with and without optimization on the hydrogen production capacities, the H_2 concentration and production rates were examined too. The sampling time interval of biogas, produced from gas collector, was 24h. The obtained results are shown on Figure3. From the data presented on Figure2 and Figure3it can be expected that the prediction should be harder in case of uncontrolled system – from one side some sensors don't return reliable data after some time, from the another side there is a fall in the hydrogen production and concentration.

In order to couple the input and output values together, the time ranges have to be unified. Since the hydrogen production to concentration rates consisted only of 19 values, the sensor measurements of time intervals are interpolated.For each 24-hours interval, new 40 points are defined using linear interpolation. To the obtained sets, a noise function (normal distributed random values with defined standard deviation value)is added, whereas the average noise is not more than 3.5% of the maximal value. Therefore for the concentration values the function with standard deviation of 1.5is used as noise. The noise for the production of hydrogen is defined using the same function but with the standard deviation of 0.1. All calculations are performed using Neural Network Toolbox from MathWorksMatlab® version 7.10 (R2010a).

3. Results and Discussion

The analyzed set consists of 721 data points. The70% of them is used for training and 15% for the test and validation setsrespectively. For each tested network new sets are chosen randomly with uniform distribution. The training is performed using Levenberg-Marquardt backpropagation method and the performance is measured using Mean Squared Error and R^2 values for the validation set.

Tested networks consist of 6 neurons in input layer, one per sensor and one for time, variable amount of neurons in the hidden layer and one output neuron for the hydrogen production or concentration. The presence of one up to ten neurons in the hidden layer is tested. The results are presented on **Figure**4 and **Figure**5 and the data can be found in the

Error! Reference source not found. and **Error! Reference source not found.**. The Mean Square Error (MSE) and the coefficient of determination (R^2) are select to assess the agreement of training and predictive grade of the NNs.

For all cases similar results behavior can be noticed. By one hidden neuron the performances are significantly worse than by two or more of them, concerning both MSE and R^2. With increasing number of neurons the error values are lowered and goodness of fit grows. There can be some fluctuations noticed (like R2 on **Figure**4 right, points 8 and 9), but the trend remains constant.

As expected, the networks for the controlled system show lower error values and better goodness of fit. The values for uncontrolled system are slightly worse, because the network had to model the production and concentration lowering after 200 hours and the noise generated by some sensors at the end of the measurement have also slightly negative impact on the results quality. Fortunately, the presence of other sensors returning correct results seems to compensate the erroneous signals from the sensor #5 and later #4. The controlled system is also more stable with adding additional neurons to the hidden layer. In the uncontrolled system the result fluctuations occur more often and although the general trend is preserved, it is more probable to get worse performance values after increasing hidden neuron count than in controlled system. Also inverse correlation of MSE and R^2 values for the controlled system is easier to spot.

It is noticeable, that networks with one neuron in the hidden layer provide already satisfying results with the goodness of fit over 0.9. Generally there is no reason in using more than 4 neurons in the hidden layer, since the both MSE and R^2 values for all tested constellations show stable behavior independent of neuron count.

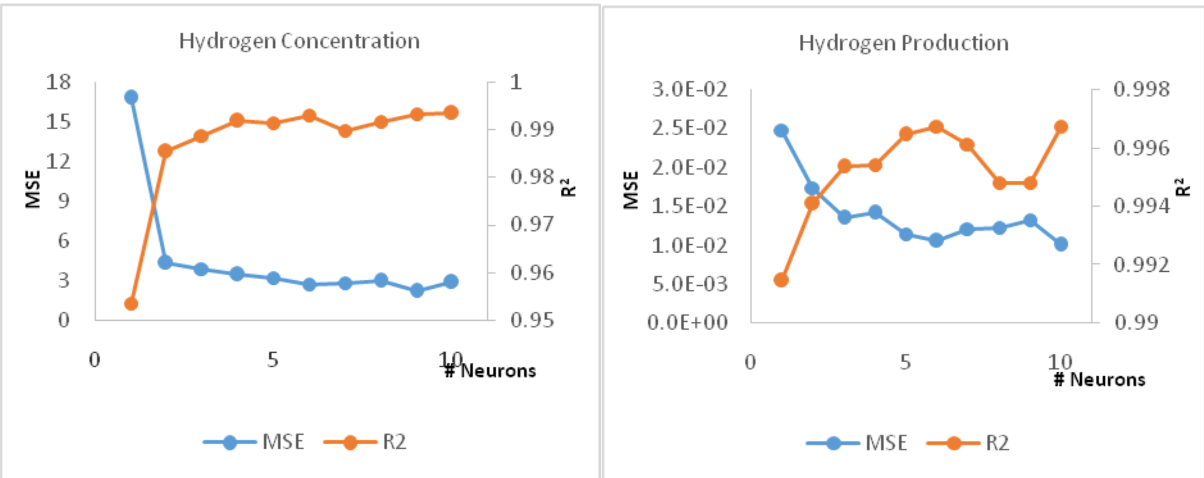

Figure4:Neural Network performance values (Mean Squared Error and R^2) for the prediction of hydrogen concentration and production rates in the controlled system.

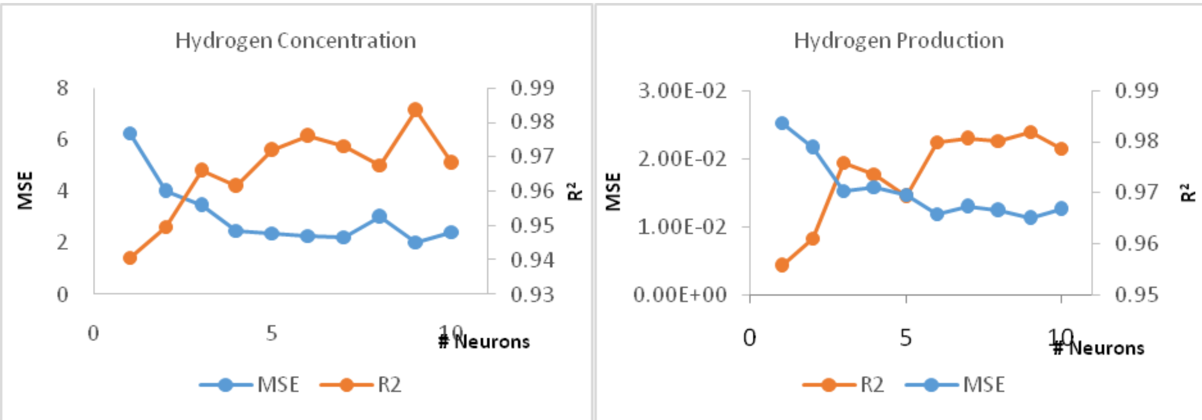

Figure5:Neural Network performance values (Mean Squared Error and R^2) for the prediction of hydrogen concentration and production rates in the uncontrolled system.

The results show clearly that it is possible to build an effective system for the prediction of the hydrogen production and concentration rates based only on the data covering biofilm thickness.

4. CONCLUSIONS AND FURTHER WORK

An artificial intelligent system for the efficient prediction of the hydrogen concentration and production is created and presented. The BPBR behavior can be accurately predicted with one hidden layer of four neurons. It can be asseverate that the experimental set up with automatic apparatus for the control of the biofilm thickness favors the development of predictive tools in a post processing of data. The advantages of data reduction by the employment of NNs can be noticed, as a predictive tool is directly and easily built from the data measurements; complementing the insight of the experimental study and offering additional perspectives for the use of collected data.

For the further work more data could be gathered, possibly from the different measurement conditions. Investigating the impact of water temperature and salinity would allow the constructing of more comprehensive system and further optimize the bio-hydrogen production process.

5. ACKNOWLEDGEMENT

We want to show our gratitude to ProfesorNianbingZhong of the Key Laboratory of Low-grade Energy Utilization Technologies and Systems in the University of Chongqing for the immeasurable collaboration in the development of the present paper.

6. DATA TABLES

Table1: Network performancevaluesforthecontrolledsystemwith different numberofhiddennodes.

# Neurons	Concentration		Production	
	MSE	R^2	MSE	R^2
10	2,85452	0,9934	1,01E-02	0,9967
9	2,1671	0,9933	1,33E-02	0,9948
8	2,94928	0,9918	1,22E-02	0,9948
7	2,82819	0,9898	1,21E-02	0,9961
6	2,71492	0,9930	1,05E-02	0,9967
5	3,195	0,9914	1,14E-02	0,9965
4	3,43247	0,9918	1,42E-02	0,9954
3	3,87323	0,9887	1,35E-02	0,9954
2	4,29401	0,9854	1,73E-02	0,9941
1	16,79553	0,9534	2,48E-02	0,9915

Table2: Network performacevaluesfortheuncontrolledsystemwith different numberofhiddennodes.

# Neurons	Concentration		Production	
	MSE	R^2	MSE	R^2
10	2,43397	0,968079	1,25E-02	0,978612
9	2,0215	0,983338	1,12E-02	0,981883
8	2,98605	0,967479	1,24E-02	0,980187
7	2,21144	0,973117	1,29E-02	0,98049
6	2,25883	0,97585	1,19E-02	0,979957
5	2,37408	0,971888	1,47E-02	0,969268
4	2,48158	0,961376	1,58E-02	0,973614
3	3,44843	0,965847	1,53E-02	0,975921
2	3,9952	0,949637	2,16E+00	0,961082
1	6,23705	0,940702	2,52E-02	0,955645

7. References

1. Nelles O., *Nonlinear system identification: from classical approaches to neural networks and fuzzy models.* Springer Publishing, New York, 2001.
2. Kilimann, H.C., Delgado A., Vogel R.F., Gänzle M.G., *A fuzzy Logic based model for the multi-stage high pressure inactivation of Lactococcuslactisssp. cremorisMG 1363.* Int J Food Microbiol., 2005. **98**: p. 89-105.
3. Bening R.M., B.T.M., Delgado A., *Initial studies of predicting flow fields with an ANN hybrid.* Advances in engineering software 2001. **32**(12): p. 895-901.
4. Diez L., Z.B.E., Kowalczyk W., Delgado A., *Investigation of multiphase flow in sequencing batch reactor (SBR) by means of hybrid methods.* Chemical Engineering Science 2007. **62**(6): p. 1803-1813.
5. Chen, T.C., H., *Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems.* IEEE Transactions on Neural Networks, 1995. **6**(4): p. 911-917.
6. Cybenko, G., *Approximation by superpositions of a sigmoidal function.* Mathematics of Control, Signals, and Systems, 1989. **3**: p. 303-314.
7. Hornik, K., *Approximation capabilities of multilayer feedforward neural networks.* Neural Networks, 1990. **4**: p. 251-257.
8. Hornik, K., *Some new results on neural network approximation.* Neural Networks, 1993. **6**: p. 1069-1072.
9. Leshno, M., Lin, V., Pinkus, A., & Shochen, S., *Multilayer feedforward networks with a polynomial activation function can approximate any function.* Neural Networks, 1993. **6**: p. 861-867.
10. Scarselli F., T., Ah Chung, *Universal Approximation Using Feedforward Neural^Networks: A Survey of Some Existing Methods and Some New Results.* Neural Networks, 1998. **11**(1): p. 15-37.
11. Yi Liao, S.-C.F., Henry L.W. Nuttle, *Relaxed conditions for radial-basis function networks to be universal approximators.* Neural Networks, 2003. **16**: p. 1019-1028.
12. Widrow, B., *30 years of adaptive neural networks: Perceptron, Madaline and Backpropagation.* IEEE Transactions on Neural Networks, 1990. **78**(9): p. 1415-1442.
13. Turkia, H., Sirén H., Pitkänen J.P., Wiebe M., Penttilä M., *Capillary electrophoresis for the monitoring of carboxylic acid production by gluconobacter oxydans.* J Chromatogr A, 2010(1217): p. 1537-1542.
14. Argun H., K.F., Kapdan I.K., Oztekin R., *Batch dark fermentation of powdered wheat starch to hydrogen gas: effects of the initial substrate and biomass concentrations.* International Journal of Hydrogen Energy, 2008. **33**: p. 6109-6115.
15. Zeng M., S.A., Roche N., *Calibration of hydrodynamic behavior and biokinetics for TOC removal modeling in biofilm reactors under different hydraulic conditions.* Bioresour Technol, 2013. **144**: p. 202-209.
16. Liao Q., Z.N., Zhu X., Huang Y., Chen R., *Enhancement of hydrogen production by optimization of biofilm growth in a photobioreactor.* International Journal of Hydrogen Energy, 2015. **40**: p. 4741-4751.