

GMDH and Neural Network Application for Modeling Vital Functions of Green Algae under Toxic Impact

Oleksandra Bulgakova¹, Volodymyr Stepashko¹, Tetayna Parshikova², Iruna Novikova²

1 International Research and Training Center of Information Technologies and Systems of the NAS and MES of Ukraine. Akademik Glushkov Prospect 40, Kyiv, 03680, Ukraine.

2 National Taras Shevchenko University of Kyiv, Volodymyrska street, Kyiv, Ukraine.

sashabulgakova@list.ru, astrid@irtc.org.ua

Abstract. *This work presents the modeling results of influence of toxic bichromate potassium on the vital functions cells of green algae using intellectual computing – Group method of data handling (MGDH) and neural network (NN) with backpropagation algorithm learning, and also regressive analysis. The results of the laboratory experiments executed at the Kyiv national university.*

In all experiments was used the toxic – bichromate potassium with concentration from 0,05 to 135 mg/l. In some experiments this toxic was used with algin acid.

The goal of our work is to get better result of forecasting influence of toxic bichromate potassium on the vital functions cells of green algae and compare results which were got using different methods.

Keywords

Group method of data handling, neural network, backpropagation algorithm, algae, bichromate potassium, forecasting

1 Introduction

Consideration the high chrome concentration into rivers, great meaning has estimate of its toxic for microscopic algae – the one of the main product of organic substance and oxygen into water's ecosystem. One of the main organisms for the estimation of toxic metal is unicellular and active green algae - *Euglena gracilis klebs* and inert – *Chlorella vulgaris*. The development of these algae can be the indicator of intensive anthropogenic water pollution. In all experiments was investigated influence of toxic – bichromate potassium ($K_2Cr_2O_7$) with concentration from 0,05 to 135 mg/l on the vital function this algae. In some experiments this toxic was used with algin acid. Also we considerate response of reaction cells *Euglena* and *Chlorella* for different state of development: logarithmical and steady.

We have made a lot of experience using intellectual computing – Group method of data handling (MGDH) and neural network (NN) with backpropagation algorithm learning, and also regressive analysis. The results of the laboratory experiments executed at the Kyiv national university. The next sections have description and result of the experiments which helps us to get better result of forecasting influence of toxic bichromate potassium on the vital functions cells of green algae and compare results which were got using different methods.

2 Theoretical Part

2.1 GMDH

GMDH is constructed by the following six procedures:

(1) Separating original data into test and control sets

Original data is separated into test and control sets. Test data is used for estimating parameters of partial descriptions which describe partial relationships of the nonlinear system. Control data is used for organizing complete description which describes complete relationships between input and output variables.

(2) Generating combinations of input variables in each layer

All combinations of two input variables (x_i, x_j) are generated in each layer.

(3) Calculating partial descriptions

For each combination, partial descriptions of the system can be calculated by applying regression analysis to testing data. Output variables of partial descriptions are called as intermediate variables.

(4) Selecting intermediate variables

L intermediate variables which give L smallest test errors calculated using control data are selected from generated intermediate variables.

(5) Iterating calculations from 2 to 5

Select L intermediate variables are set to input variables of the next layer and calculations from procedure 2 to 5 are iterated. The multilayered architecture is organized.

(6) Stopping multilayered iterative calculation

When errors of test data in each layer stop decreasing, iterative calculation is terminated. Finally, complete description of the nonlinear system is constructed by partial descriptions generated in each layer.

2.2 Neural network with backpropagation learning algorithm

Backpropagation is the most widely applied learning algorithm for neural networks. It learns the weights for a multilayer network, given a network with a fixed set of weights and interconnections. Backpropagation employs gradient descent to minimize the squared error between the networks *output values* and *desired values* for those outputs. The goal of gradient descent learning is to minimize the sum of squared errors by propagating error signals backward through the network architecture upon the presentation of training samples from the training set. These error signals are used to calculate the *weight* updates which represent the knowledge learnt in the network.

3. Result and discussion

Consideration the high chrome concentration into rivers, great meaning has estimate of its toxic for microscopic algae – the one of the main product of organic substance and oxygen into water's ecosystem. One of the main organisms for the estimation of toxic metal is unicellular and active green algae - *Euglena gracilis* klebs and inert – *Chlorella vulgaris*. The development of these algae can be the indicator of intensive anthropogenic water pollution.

In all experiments was investigated influence of toxic – bichromate potassium ($K_2Cr_2O_7$) with concentration from 0,05 to 135 mg/l on the vital function this algae. In some experiments this toxic was used with algin acid. Also we considerate response of reaction cells *Euglena* and *Chlorella* for different state of development: logarithmical and steady.

3.1 GMDH

We describe some experiments which were conducted.

Experiment 1.

We ran our experiments by using 60% of the data for testing and 30% for control and the remaining 10% for forecast. On fig.1 you can see real (y) and prognosis (y_{pred}) data which were modeling by GMDH.

1-a: *Influence of chrome (6v) on concentration (from 0,05 to 135 mg/l) on toxic – bichromate potassium ($K_2Cr_2O_7$) of *Euglena gracilis* (through 1,4,7 days after adding).* This data were described by next equation:

$y_{pred} = 411,187 - 3,234 \cdot x_1 + 74,783 \cdot x_2 + 0,028 \cdot x_1^2 - 11,976 \cdot x_2^2 - 0,329 \cdot x_1 \cdot x_2$, where x_1 – concentration of chrome, x_2 – the day number

1-c: *Influence of toxic – bichromate potassium on activity cells of *Euglena gracilis* (through 1,4,7 days).* This data were described by next equation:

$y_{pred} = 25,069 + 0,168 \cdot x_1 + 0,095 \cdot x_2 - 0,0019 \cdot x_1^2 + 0,1493 \cdot x_2^2 - 0,0183 \cdot x_1 \cdot x_2$, where x_1 – concentration of chrome, x_2 – the day number

1-d: Influence of toxic – bichromate potassium on change of energy movement cells of *Euglena gracilis* (through 1,4,7 days). This data were described by next equation:

$y_{pred} = 2,463 + 0,0054 \cdot x_1 + 0,28 \cdot x_2 - 0,0001 \cdot x_1^2 + 0,003 \cdot x_2^2 - 0,0037 \cdot x_1 \cdot x_2$, where x_1 – concentration of chrome, x_2 – the day number

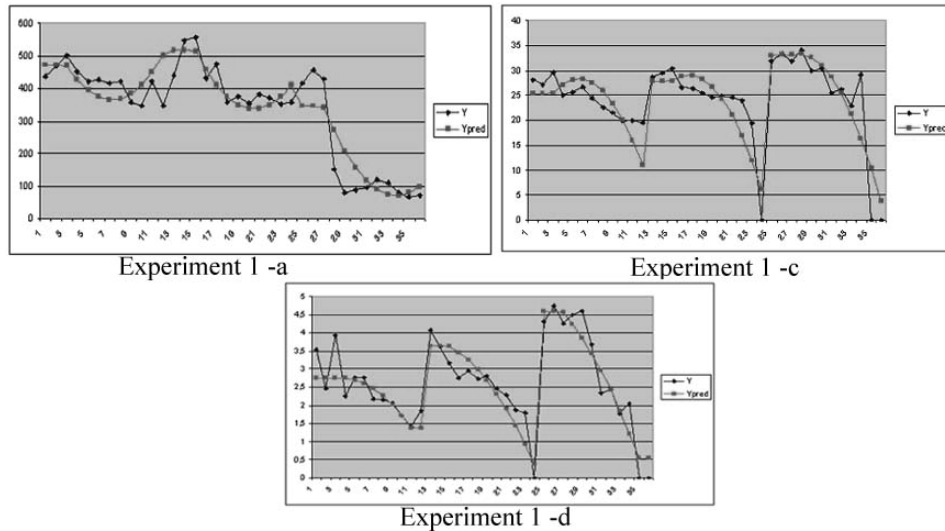


fig.1. Experiment 1

Experiment 2, 3. Explore reaction response cells of *Euglena gracilis* on the difference phases of development.

On fig.1 you can see real (y) and prognosis (y_{pred}) data which were modeling by GMDH.

2-c: Influence of bichromate potassium ($K_2Cr_2O_7$) on amount cells of *Euglena gracili*) for logarithmical phase. This data were described by next equation:

$$y_{pred} = 2287,21 - 45,21x + 0,533x^2 - 0,002x^3, \text{ where } x - \text{concentration.}$$

3-b: Influence of toxic – bichromate potassium (b_1) and – bichromate potassium with algin acid (b_2) on change concentration for steady state. This data were described by next equation:

$$(b_1): y_{pred} = 178,98 - 3,13x + 0,04x^2 - 0,0002x^3,$$

$$(b_2): y_{pred} = 175,37 - 3,71x + 0,04x^2 - 0,00015x^3 \text{ where } x - \text{concentration.}$$

On fig.2 you can see real (y) and prognosis (y_{pred}) data which were modeling by GMDH.

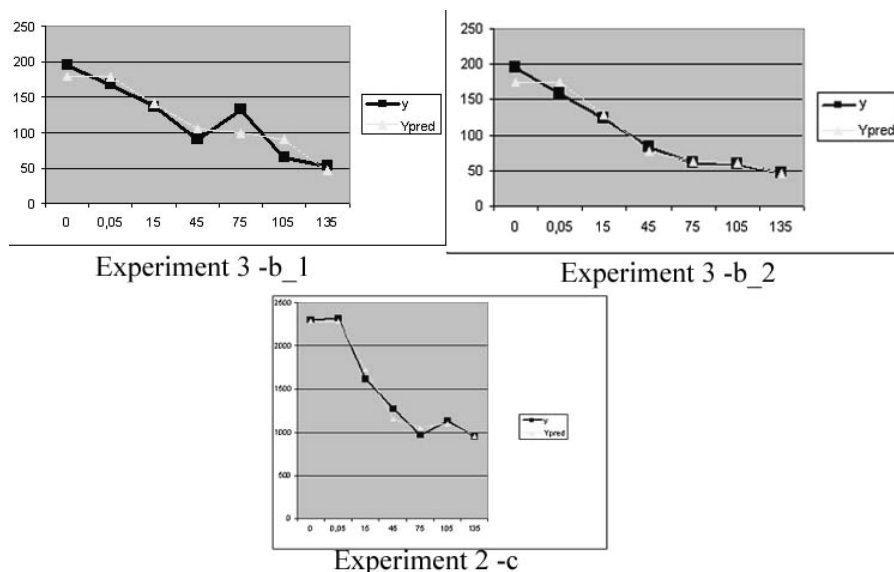


Fig.2. Experiment 2, 3.

Experiment 4.

We ran our experiments by using 60% of the data for testing and 30% for control and the remaining 10% for forecast

4-a: Influence of bichromate potassium ($K_2Cr_2O_7$) on concentration for *Chlorella vulgaris* (through 1,4,7 days after adding). This data were described by next equation:

$y_{pred} = 8,202 - 0,28 \cdot x_1 + 19,53 \cdot x_2 + 0,0018 \cdot x_1^2 - 2,137 \cdot x_2^2 - 0,34 \cdot x_1 \cdot x_2$, where x_1 – concentration of chrome, x_2 – the day number

4-c: Change speed of efflorescence chlorophyll for *Chlorella vulgaris* (through 1,4,7 days). This data were described by next equation:

$y_{pred} = 85,17 - 1,75 \cdot x_1 + 1,057 \cdot x_2 + 0,00912 \cdot x_1^2 - 0,21 \cdot x_2^2 + 0,079 \cdot x_1 \cdot x_2$, where x_1 – concentration of chrome, x_2 – the day number

On fig.3 you can see real (y) and prognosis (y_{pred}) data which were modeling by GMDH.

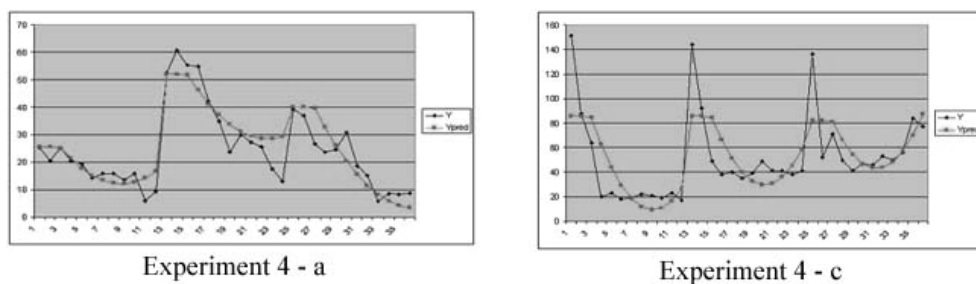


Fig.3. Experiment 4.

3.2 Compare GMDH and NN. Interpolation model.

We ran our experiments by using 90% of the data for testing (testing and control) and the remaining 10% for forecast. The first part was to determine the optimal values for the number of hidden neurons and the learning rate. The neurons in the network had a sigmoidal discriminant function and all networks were trained using the standard quadratic error function.

For compare GMDH and NN we used experiment 1-a. On fig.4 (GMDH, NN) you can see real (y) and prognosis (y_{pred}) data which were modeling by GMDH and NN. For more exact description of model, we build dependence of output variable which allows reproducing absent information for the intervals of time and concentrations. For this we build dependence output variables from time:

$$y_{int}(t) = a_0 + a_1 \cdot x_1 + a_2 x_1^2 + a_3 \cdot x_1^3 + a_4 \cdot x_1^4, \text{ where } x_1 \text{ – concentration and coefficients are equals:}$$

$$a_n = b_{0n} + b_{1n} \cdot t + b_{2n} \cdot t^2, \quad t = 1,4,7; \quad n = 1,2,3,4;$$

On this basis, we got next model:

$$y_{int}(t) = (425,93 - 48,1 \cdot t - 6,59 \cdot t^2) + (-449,37 + 478,37 \cdot t - 112,23 \cdot t^2) \cdot \frac{x_1}{100} +$$

$$+ (1185,9 - 1626,8 \cdot t + 334,17 \cdot t^2) \cdot \left(\frac{x_1}{100}\right)^2 + (-1269,4 + 1780,6 \cdot t - 344,75 \cdot t^2) \cdot \left(\frac{x_1}{100}\right)^3 +$$

$$+ (437,18 - 614,65 \cdot t + 115,24 \cdot t^2) \cdot \left(\frac{x_1}{100}\right)^4$$

where $(x_1/100)$ – norm of x_1 . On fig.4 (Int – 1,4,7) you can see real (y) and prognosis (y_{int}) data which were modeling by Interpolation model. On fig.4 (Int – 1,2,3,4,5,6,7) you can see the family of curves for different concentrations (for 1,2,3,4,5,6,7 days).

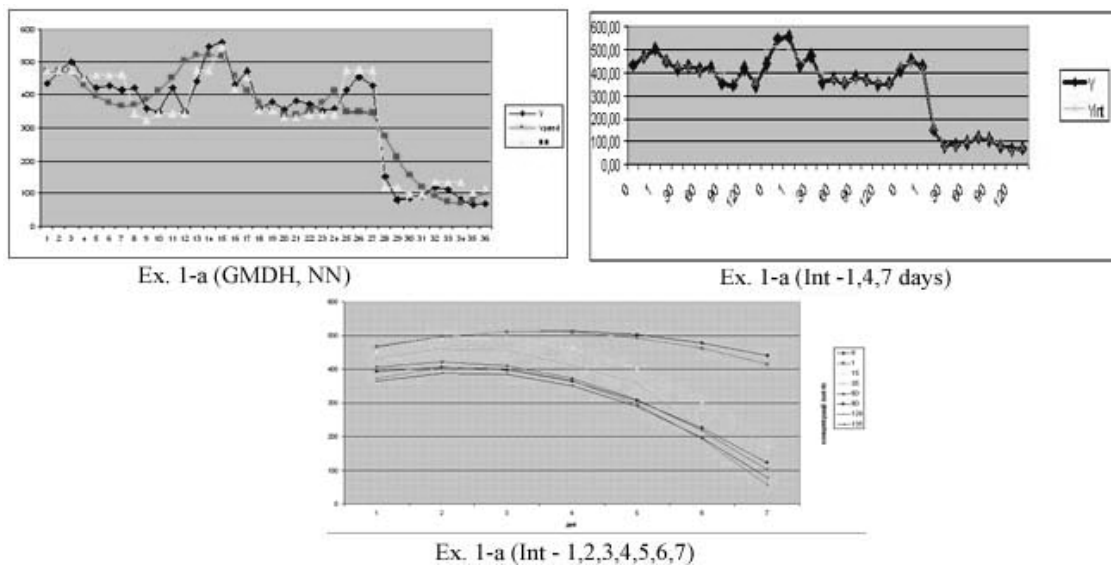


Fig.4.

Conclusion

We have made a lot of experience using intellectual computing – Group method of data handling (GMDH) and neural network (NN) with backpropagation algorithm learning, and also regressive analysis. The results of the laboratory experiments executed at the Kyiv national university.

For the best prognosis, we developed an interpolation model and build dependence of output variables which allows to reproduce absent information for the intervals of time and concentrations. Exactness of this model made 97%, that on 12% more than earlier findings by GMDH for these experimental information.

References

- [1] Buerge J., Hug J. Influence of Mineral Surfaces on Chromium (VI) reduction by Iron (II) // Environ. Sci. Technol. – 1999. – 33. – P. 4285 – 4291.
- [2] Stepashko V.S. GMDH algorithms – the basic of the automatization process of modeling on experimental data// Automatics-1998.-№ 4.- c. 44-55.
- [3] Robert Callan. Essence of Artificial Intelligence. Prentice Hall (1997).