ISSN 0026-2617, Microbiology, 2007, Vol. 76, No. 2, pp. 229–235. © Pleiades Publishing, Ltd., 2007.
Original Russian Text © T.I. Lobova, Yu.P. Lankin, L.Yu. Popova, 2007, published in Mikrobiologiya, 2007, Vol. 76, No. 2, pp. 263–270.

EXPERIMENTAL ARTICLES

Assessing the Anthropogenic Impact on Lake Shira from Antibiotic Resistance of Heterotrophic Bacteria by Neural Networks Methods

T. I. Lobova[a,1], **Yu. P. Lankin**[b], **and L. Yu. Popova**[c]

[a] *Territory-Oriented Informational Systems, Ltd. (Institute of Computational Modelling), Siberian Division, Russian Academy of Sciences, Krasnoyarsk*

[b] *Institute of Biophysics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*

[c] *Krasnoyarsk Research Center, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*

Received March 13, 2006; in final form, August 18, 2006

Abstract—A general approach to assessing the anthropogenic impact on lake ecosystems is proposed and exemplified for the case of Lake Shira (Republic of Khakasia, Russia). The impact strength is estimated by applying neural network–based methods to samples of data on interdependent marking features of autochthonous and allochthonous bacteria isolated from the lake in 1997–2001. The proposed combination of analysis methods makes it possible to determine the state of an ecosystem from both small- and large-size samples of data having complex interrelations.

DOI: 10.1134/S0026261707020154

Key words: heterotrophic bacteria, antibiotic resistance, neural networks, ecological models.

The increasing anthropogenic impact on natural ecosystems makes their preservation a topical problem that can hardly be solved unless methods of monitoring their condition are developed. In this work, we present results of processing experimental data to assess the condition of an aquatic ecosystem of Lake Shira (Republic of Khakasia, Russia), used as an example.

The highly mineralized Lake Shira is one of the larger water bodies in the Russian Federation and is well-known for its unique healing power. Being a highly popular tourist site and a health resort, this lake needs to be monitored for anthropogenic impact on its ecosystem. In our previous studies, it was established that multiple antibiotic resistance exhibited by heterotrophic bacteria from this lake can serve as a marker for anthropogenic impact on its ecosystem and be used to monitor its ecological condition [1, 2].

An important issue in monitoring anthropogenic impact on natural ecosystems is the choice of methods to assess the strength of this impact. In most cases, standard mathematical methods are employed to formulate and analyze models that can range from simple differential equations to fairly complex sets of equations. Even with a large dataset, however, a researcher may fail to establish a link between features of the object at hand and the environmental factors giving rise to these features. A solution to this problem can be found in employing a neural network-based approach to analyze experimental data.

Neural networks represent a thriving and highly promising field of mathematics and computer science. The term "neural network" was borrowed from biology: the structure and operational rules of artificial neural networks are similar to those of the neural networks of the brain. An important advantage of the neural network approach consists in its ability to handle some extremely complex problems that are impossible or very difficult to solve by conventional mathematical methods [3–8].

Under conventional approaches, an algorithm to solve the problem that a researcher has to come up with will most likely consist of a long sequence of mathematical operations, which will inevitably include a multitude of errors that are hard to eliminate. Unlike conventional mathematical methods, approaches based on neural networks are free from this shortcoming because the problem is solved by training a neural network. In many instances, neural networks prove a more efficient instrument for tackling complex problems such as forecasting, classification, approximating com-

[1] Corresponding author. E-mail: yanalobova@mail.ru

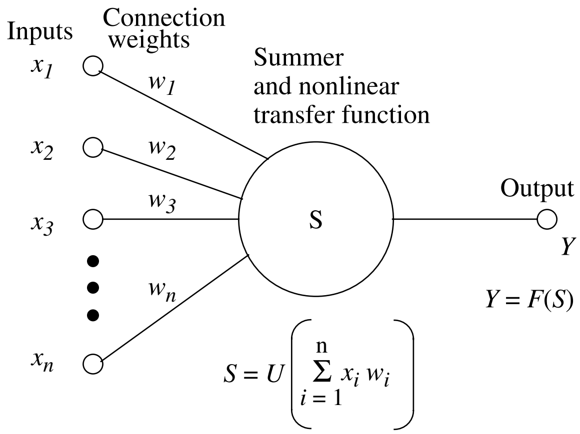

Fig. 1. Diagram of an artificial neuron as conceptualized in artificial neural networks.

plex functions, image recognition, etc. Neural networks can also be applied to great advantage in information processing and modeling of biological processes.

The goal of this work consisted in developing a neural network-based method for estimating the level of anthropogenic impact on Lake Shira from data samples with complex interdependencies.

MATERIALS AND METHODS

The objects of study were heterotrophic bacteria isolated from various zones of the brackish water Lake Shira in 1997–2001. Samples of water were plated onto agarized medium M9 modified by the addition of peptone and containing (per liter of distilled water) Na_2HPO_4, 6 g; KH_2PO_4, 3 g; NaCl, 3 g; NH_4Cl, 1 g; peptone, 5 g; and agar–agar, 20 g. Upon autoclaving, $MgSO_4$ (20% solution), 1 ml; $CaCl_2$ (0.5% solution), 1 ml; and glucose (20% solution), 10 ml were sterilely added to the medium [9]. In order to isolate nonhalophilic and halotolerant heterotrophic bacteria, medium M9 with peptone was supplemented with sodium chloride in concentrations of 0.5, 50, and 100 g/l.

Determining antibiotic resistance of bacteria. The antibiotic resistance of the isolated bacteria was determined on agarized medium M9 with peptone containing two antibiotics with different mechanism of action: ampicillin (inhibiting cell wall synthesis) and kanamycin (inhibiting protein synthesis). These antibiotics were added to the medium upon autoclaving at concentrations of 50 μg/ml. The antibiotic resistance in heterotrophic bacteria was determined from the numbers of bacteria (1) resistant to ampicillin, (2) resistant to kanamycin, (3) resistant both to ampicillin and kanamycin, and (4) susceptible to antibiotics.

Brief description of neural networks. Most often, artificial neural networks are designed in the form of computer programs that implement certain mathematical algorithms. Let us now consider a simple example of a neural network. Mathematically, artificial neural networks are composed of artificial neurons (Fig. 1) described by algebraic equations. Each neuron in the network has several inputs (x_1–x_n), whose number can vary depending on the specifics of the task, and only one output *Y*. The term *input of the neural network* refers to the inputs of the constituting neurons that receive the input data. The term *output* designates outputs of the neurons that represent the obtained results. Mathematically, the inputs are characterized by numeric factors (w_1–w_n) in the algebraic equations (Fig. 1), while outputs are solutions to these equations. The values of experimental data fed to the network input are multiplied by the corresponding weights and then summed up. In some neural algorithms (i.e., mathematical algorithms describing training and operation of neural networks), the obtained net inputs are further modified by a simple nonlinear transfer function, the same for all neurons. The neuron's output value (a solution to the algebraic equation) is fed as input to several other neurons, or represents the output of the neural network. The values obtained as the neural network's outputs are the obtained solutions to the given problem. In order to solve the problem, the network first has to be trained.

The training of an artificial neural network to solve a particular problem consists in adjusting the scalar weights of its interneuron connections by a special algorithm. Upon the completion of training, the neural network is ready for use. The result of training is captured in the weight matrix of neuron interconnections.

In order to train a neural network, one must have a comprehensive sample of source data, reflecting general patterns of their change. A trained neural network has a high processing speed and, if needed, can be further trained when new experimental data become available.

Network training algorithms. Two types of neural network training algorithms were used in this work. The first one, a dual functioning algorithm [11], belonging to the class of supervised training algorithms, is a generalization of the well-known error back-propagation algorithm [12] implemented in the MODEL program [6]. In terms of its capacity, this algorithm [11, 12] is second only to the self-adaptation algorithm [13]. To use the dual functioning algorithm, one needs to compile a table of "examples" of problem solutions. Each row of such a table contains an instance of experimental data and the required response of the neural network. The second algorithm of neural network training that we used was Kohonen's self-organizing feature maps [7]. This algorithm belongs to the group of unsupervised training algorithms and allows experimental data to be classified without providing any classification examples.

RESULTS AND DISCUSSION

Preparation of experimental data for computer analysis. The specific features associated with ecological specialization of autochthonous and allochthonous bacteria isolated from Lake Shira were identified in our previous studies. Representatives of autochthonous bacteria were shown to be able to grow at temperatures as low as 5°C and at concentrations of sodium chloride as high as 10%. At the same time, no traits of either psychro- or halotolerance were observed in the allochthonous bacteria, which are characterized by elevated antibiotic resistance and occur in the parts of the aquatic ecosystem exposed to stronger anthropogenic influence [14]. In this work, the anthropogenic impact on Lake Shira was estimated in terms of the available experimental evidence on halotolerance and antibiotic resistance as marker features for autochthonous and allochthonous bacteria. The corresponding bacteria were designated as follows: nonhalophilic bacteria, G1; halophilic bacteria, G2; moderately halotolerant bacteria, G3; bacteria resistant to ampicillin, A1; bacteria resistant to kanamycin, A2; bacteria with multiple antibiotic resistance, A3; and bacteria sensitive to antibiotics, A4.

In this work, three possible conditions of Lake Shira were considered: (1) strong anthropogenic influence characterized by the predominance of heterotrophic bacteria with multiple antibiotic resistance; (2) moderate anthropogenic influence, when at least 50% of the total quantity of bacteria isolated from samples of water taken on any date exhibit multiple resistance to antibiotics; and (3) low anthropogenic influence, when the lake is dominated by bacteria sensitive to antibiotics. For ease of discussion, these three states of the lake are designated as Red, Yellow, and Green.

The vast body of experimental evidence accumulated over many years of field-work on Lake Shira is hard to classify according to the condition of its ecosystem; and standard mathematical methods failed to expose the corresponding patterns in the experimental data. To reveal the intrinsic interdependencies in the available data, a tentative analysis of the data structure was undertaken. As a first step in preparing the data for computer analysis, the parameter vectors characterizing the state of Lake Shira were assigned by an expert in microbiology to one of the three above-mentioned ecological conditions: Red, Yellow, or Green.

Preliminary analysis of experimental data by mathematical and graphical methods. This stage of work consisted in analyzing the parameters introduced above with the goal of identifying the most significant ones for the problem of classifying the ecological conditions of Lake Shira. This was accomplished by employing methods of statistical estimations, by data processing using neural networks, and by graphical data analysis. By combining these methods, the most significant parameters, sufficient for solving the problem, were identified as A1, A3, and A4. A useful technique for determining whether a set of input vectors can be well partitioned into classes is data visualization. In our case, the two most significant parameters for classification, among the three under consideration, were found to be A3 and A4. Such a reduction of a 3D problem to a 2D problem simplifies the subsequent data analysis and makes it unnecessary to employ more intricate methods. With just two parameters, the available dataset can be conveniently graphed in the plane (Fig. 2).

It can be seen from Fig. 2 that the input vectors fall into three compact regions corresponding to the three introduced classes (Red, Yellow, and Green). This partition into regions basically accomplishes the desired classification. Still, as one can see from Fig. 2, the obtained regions do overlap, and so in some cases the ecological conditions cannot be uniquely determined. Fortunately, we can accomplish this task by taking into account the third vector component A1. However, even in the 3D space, there is no simple way of breaking the set of input vectors into the three considered classes. For example, an attempt was made to do this by applying correlation analysis to input vectors represented by components A1, A3, and A4. Dividing input vectors into classes by simple (linear) decision rules requires that the pairwise correlation coefficients of vectors belonging to one class be high and those belonging to different classes be low. It was found, however, that there existed pairs of vectors from different classes with correlation coefficients much higher than those of a large number of vector pairs from one class. This fact is readily seen from a fragment of the correlation table given in Table 1, where fairly high correlation coefficients for vectors of one class are printed in boldface, quite small correlation coefficients for vectors of one class are printed in italics, and extremely high correlation coefficients for vectors from different classes are printed in grey. The above-mentioned fact makes it impossible to linearly separate vectors into the given classes of ecological conditions. The indices of the corresponding vectors in the input dataset are listed in the left column and the upper row of Table 1.

These inconsistencies, revealed by correlation analysis, are well illustrated by graphs in Fig. 3. An example of two input vectors belonging to one class is given in Fig. 3a and an example of vectors belonging to different classes is shown in Fig. 3b. The first letter in the designation of each graph refers to the base class of ecological conditions (Red, Yellow, or Green), with the number next to it indicating the subclass within the base class (the numbers in brackets are the ordinal numbers of the corresponding vectors in the input dataset).

Examination of the obtained results showed that input data could not be partitioned into the given ecological condition classes unless intrinsic properties of the input data are revealed and taken in to account. The main problem consists in classifying vectors that fall into marginal zones of the classes (Fig. 2). In these

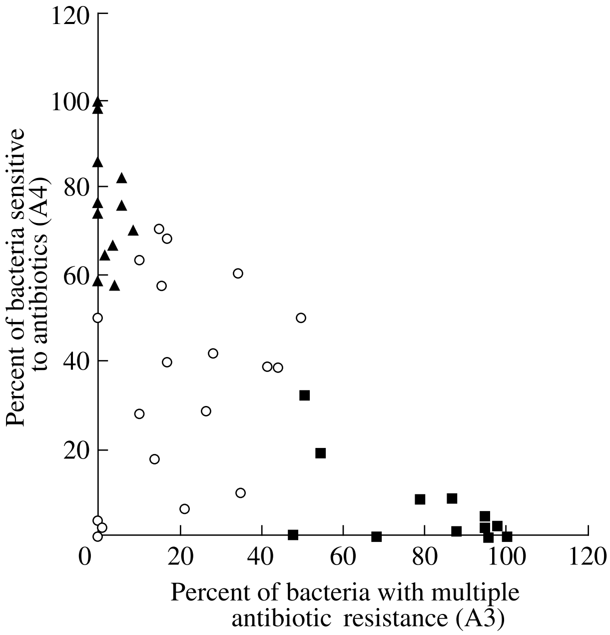

Fig. 2. 2D diagram showing grouping of input vectors in terms of their two selected components. The base classes are Red (■), Yellow (○), and Green (▲).

zones, the similarity of vectors from different classes is much higher (Fig. 3b) than that of vectors from the same class lying further away from this zone (Figs. 2 and 3b).

It is worth noting that, in such cases, finding a marginal zone for the two classes is not at all a trivial task. The available data points may well fail to form uniform groups, and, for example, may aggregate in several separate regions, where vectors from different classes could occur. The problem of classification in such circumstances is mathematically quite difficult. One approach to classifying data with such a complex structure consists in splitting the given classes (Red, Yellow, and Green) into smaller subclasses such that vectors within these subclasses have greater similarity than vectors of neighboring subclasses from a different class.

Table 1. A fragment of the table of pairwise correlations of data vectors to be classified

	4	5	6	7	8	9
26	...	...	...	...	...	...
27	...	*0.1581*	**0.9861**	*0.3491*	*0.567*	...
28	...	0.9758	*0.5195*	**0.9997**	*0.556*	...
29	...	0.7304	**0.8809**	**0.8504**	*0.058*	...
30	...	**0.9757**	*0.1046*	0.9136	0.858	...
31	...	...	...	...	...	...

Note: The numbers in the first column and in the first row refer to the vector's ordinal numbers in the dataset. For designations, see text.

Methods of unsupervised data classification provide some useful tools for dealing with such problems. One option is to use neural networks with Kohonen feature map architecture [7]. In our case, this method failed to yield a definitive solution to the problem at hand because of the complex structure of the input data. It can still be used to greatly reduce the complexity of the preliminary data classification and of identifying marginal zones.

At this stage of analysis, it is important to determine the minimal number of classes one needs to solve the problem. In our case, the whole of the source dataset (a collection of vectors) was assigned to seven subclasses by the Kohonen network. Some of the obtained subclasses coincided with the three base classes (Red, Yellow, and Green), while the rest of subclasses identified by the Kohonen network contained mixed data. The falling of vectors from two different classes within one subclass is an indication that such a subclass needs to be further split into two individual subclasses, each belonging to its own base class. The final decision regarding the partitioning of vectors into subclasses was made by taking into account the results of correlation analysis (Table 1) and graphical (Figs. 2–4) examination of the input data. The overall number of the obtained subclasses was eleven, composed of two red, six yellow, and three green subclasses. An example of assignment of close vectors to two subclasses is shown in Fig. 4.

Methods employed to solve the problem. Using the obtained partitioning, a training dataset was compiled to train the neural network by the algorithm back-propagation error [11, 12] as implemented in the computer program MODEL [6]. This program was specifically designed to solve difficult nonlinear problems. By using this algorithm, the 11 above-mentioned subclasses were identified in the input collection of 3D vectors.

A fragment of this training dataset is presented in Table 2. Each row of the table contains ten zeros and only one unity, indicating the subclass (within the Red, Yellow, or Green classes) selected by the neural network as a result of input data classification. Each subclass corresponds to one of the outputs of the neural network. The third header row of Table 2 lists the subclass number within the corresponding base class. The values A1, A3, and A4 are fed into the neural network as its inputs and the obtained output values are the subclass numbers assigned by the network. As previously noted, the value of “1” obtained on one of the eleven outputs refers the input vector to the subclass this output is associated with. Having trained the neural network, we had an expert system able to assess the ecological conditions in Lake Shira. Its results proved to be very close to the judgments made by an expert microbiologist.

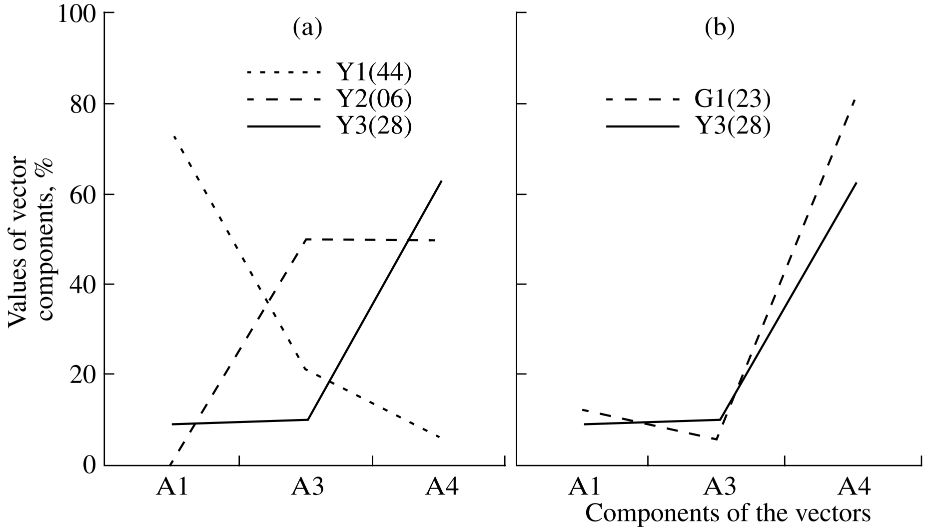

Fig. 3. Examples of (a) dissimilarity of vectors belonging to one class and (b) similarity of vectors from different classes.

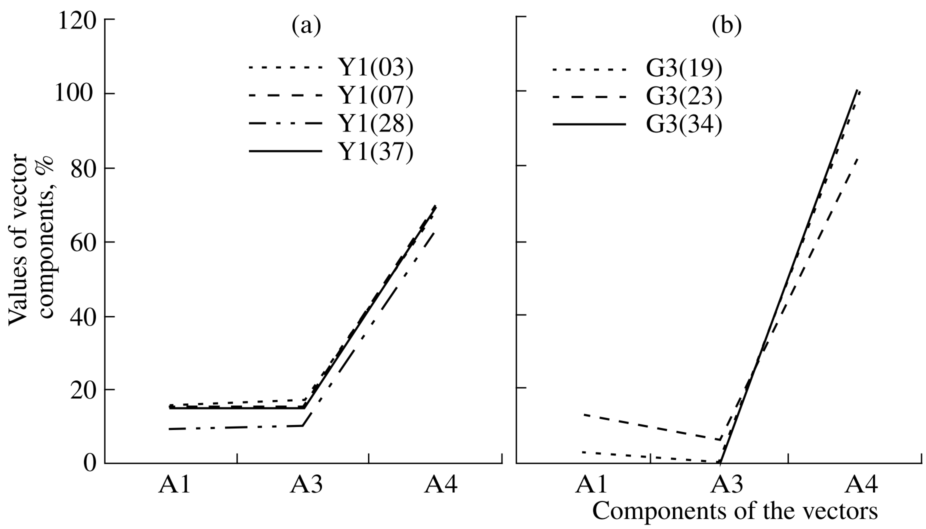

Fig. 4. Examples of similarity of vectors from two different classes: (a) Yellow and (b) Green.

Shown in Fig. 5 is the variation of ecological conditions from 1997 through 2001 for each of the four stations in Lake Shira. The ordinate axis in each graph indicates the base class number: 1 stands for Green, 2 is Yellow, and 3 is Red. The heights of the bars in these graphs reflect the strength of the anthropogenic impact on the lake.

In this paper, a method employing neural networks for unsupervised processing of experimental data with complex intrinsic structure was developed and shown to work on experimental data relating to marker features of heterotrophic bacteria isolated from Lake Shira. This method can be used by a wide group of experts in the field of environmental monitoring.

CONCLUSIONS

The results presented in this work appear to be important from three points of view. From the fundamental point of view, the main result consists in the proposed method to assess the state of objects as complex as lake ecosystems on the basis of just a few parameters (in our case, these were A1, A3, and A4) identified by mathematical analysis of the data.

Methodologically, the valuable result of this work consists in the possibility to analyze and process experimental data with complex interdependencies, especially in small-sized samples. It is noteworthy that the combination of data processing methods employed in this study not only made it possible to

Table 2. A fragment of the dataset compiled to train the neural network

no.	Outputs											Inputs		
	Red		Yellow						Green			Antibiotic resistance		
	1	2	1	2	3	4	5	6	1	2	3	A1	A3	A4
1	0	0	1	0	0	0	0	0	0	0	0	68.2	14	17.8
2	0	0	0	1	0	0	0	0	0	0	0	19.8	41.5	38.7
3	0	0	0	0	1	0	0	0	0	0	0	15.4	16.8	67.8
4	0	0	1	0	0	0	0	0	0	0	0	40.7	26.4	28.5
5	0	0	0	0	0	0	0	0	0	1	0	21.3	8.5	70.2
6	0	0	0	1	0	0	0	0	0	0	0	0	50	50
7	0	0	0	0	1	0	0	0	0	0	0	15	15	70
8	0	1	0	0	0	0	0	0	0	0	0	22.5	55	19
9	1	0	0	0	0	0	0	0	0	0	0	3	95.3	1.7
10	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Note: The numbers in the first column indicate the ordinal number of the corresponding vector in the dataset.

assess the ecological conditions of the given ecosystem but can also be used in assessing conditions of other ecosystems from samples of data with intricate interdependencies.

From the practical point of view, an attempt was made to develop a general approach to anthropogenic impact assessment for objects as complex and particular as the ecosystem of Lake Shira. The proposed

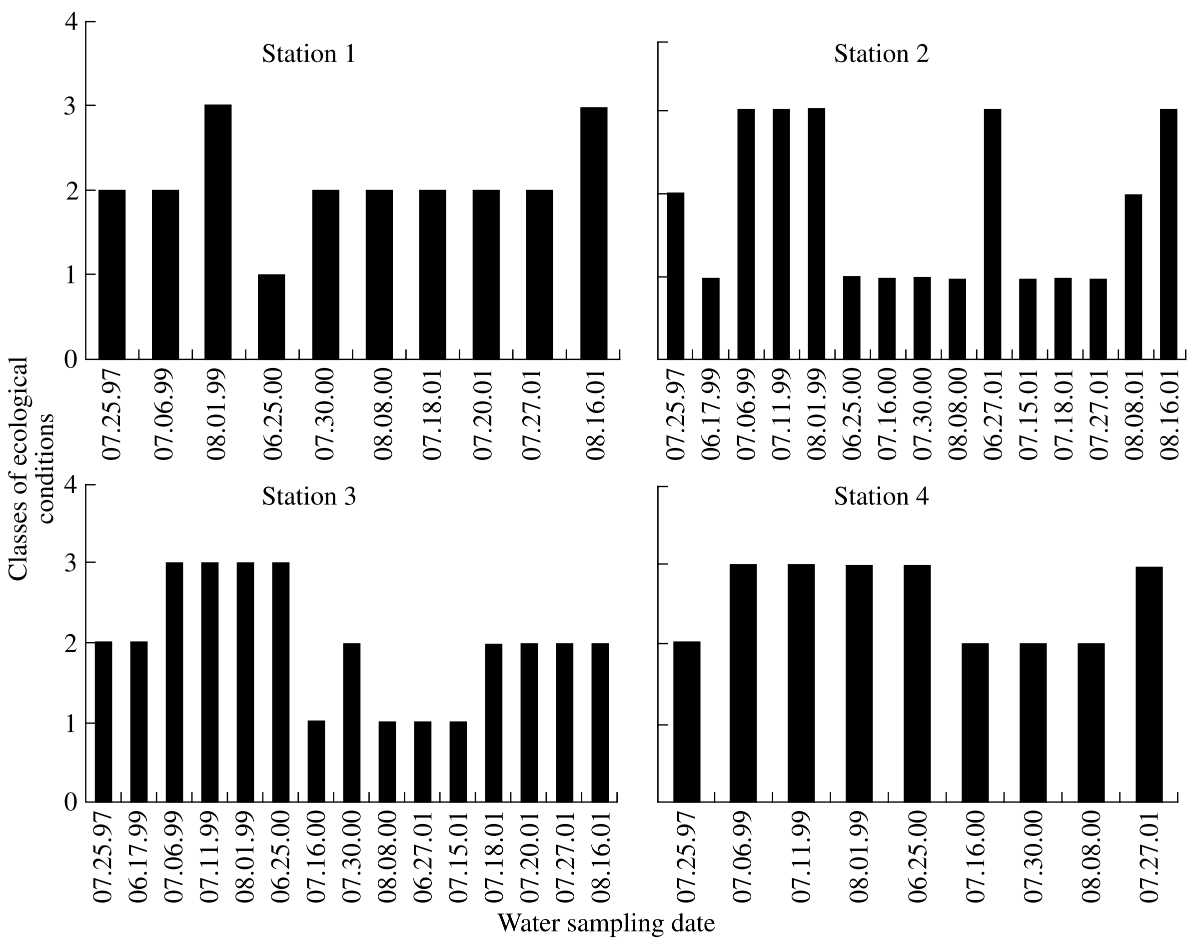

Fig. 5. Variation of ecological conditions at each station as a function of time.

method makes possible express analysis of incoming data and does not call for microbiological expertise to assess the conditions of aquatic ecosystems. The trained neural network can be used to classify experimental data that will be obtained in future studies of anthropogenic influence on the given ecosystem. The expert system we developed to assess the ecological conditions of the lake is easy to use and does not call for special mathematical background on the part of the experimentalist.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 04-05-64188.

REFERENCES

1. Lobova, T.I., Maksimova, E.E., Popova, L.Yu., and Pechurkin, N.S., Geographical and Seasonal Distribution of Multiple Antibiotic Resistance of Heterotrophic Bacteria of Shira Lake, *Aquat. Ecol.*, 2002, vol. 36, no. 2, pp. 299–307.
2. Lobova, T.I., Barchatov, Yu.V., and Popova, L.Yu., Antibiotic Resistance of Heterotrophic Bacteria in Shira Lake: Natural and Anthropogenic Impacts, *Aquat. Microb. Ecol.*, 2002, vol. 30, no. 1, pp. 11–18.
3. Kallan, R., *Osnovnye kontseptsii neironnykh setei* (Neural Networks: Major Concepts), Moscow: Vil'yams, 2003.
4. Takatsuka, M. and Jarvis, R.A., Encoding 3D Structural Information Using Multiple Self-Organising Feature Maps, *Image Vision Computing*, 2001, no. 3, pp. 99–118.
5. Bayro-Corrochano, E. and Vallejo, R., Geometric Processing and Neurocomputing for Pattern Recognition and Pose Estimation, *Pattern Recognition*, 2003, vol. 36, no. 12, pp. 2909–2926.
6. Okhonin, V., Okhonin, S., Ils, A., and Ilegemres, M., Neural Network Based Approach to the Evaluation of Degradation Lifetime, *Neural Net. World*, 2001, vol. 11, no. 2, pp. 145–151.
7. Kohonen, T., *Self-Organizing Maps. Springer Series in Information Sciences. Third Extended Edition*, Berlin: Springer, 2001, vol. 30, p. 501.
8. Gorban', A.N. and Rossiev, D.A., *Neironnye seti na personal'nom komp'yutere* (Neural Networks on a Personal Computer), Novosibirsk: Nauka, 1996.
9. Miller, J.H., *Experiments in Molecular Genetics*, New York: Gold Spring Harbor Lab., 1972.
10. *Microbial Life in Extreme Environments*, Kushner, D.J., Ed., London: Academic, 1978.
11. Bartsev, S.I. and Okhonin, V.A., Variation Principle and the Algorithm of Dual Functioning: Examples of Applications, *Neurocomputers and Attention. II: Connectionism and Neurocomputers*, Manchester Univ. Press, 1991, pp. 453–458.
12. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., Learning Representations by Back-Propogating Errors, *Nature*, 1986, vol. 323, pp. 533–536.
13. Lankin, J.P. and Baskanova, T.F., Algorithms of Self-Adaptation for Atmospheric Model Designing, *Proc. SPIE–Int. Soc. Opt. Eng.*, 2004, vol. 5397, pp. 260–270.
14. Lobova, T.I., Listova, L.V., and Popova, L.Yu., Distribution of Heterotrophic Bacteria in Lake Shira, *Mikrobiologiya*, 2004, vol. 73, no. 1, pp. 105–110 [*Microbiology* (Engl. Transl.), vol. 73, no. 1, pp. 89–93.