Application of artificial neural networks in solving water management problems

E V Venitsianov and M N Skonechnii
Institute of Water Problems, Russian Academy of Sciences, Moscow
eugeny.venitsianov@gmail.ru

Abstract. Overview of the main directions of using artificial neural networks for water monitoring and quality problems: optimization of monitoring networks, modelling of the state of aquatic ecosystems and diffuse pollution. Conclusions about the prospects and functional capabilities of artificial neural networks in solving water management problems.

1. Introduction
Modelling of the hydrochemical and hydrodynamic state of water bodies is the basis for effective water management. One of the modelling tools is artificial neural networks (ANN), i.e. computer systems that simulate the work of the human brain. The structure of the ANN consists of a number of artificial neurons connected to a network. Artificial neurons are adders of input signals, which, after conversion, are received at the input of a certain nonlinear function, i.e., the input of a certain nonlinear function. ANN allows you to solve problems with non-linear relationships of variables and identify hidden patterns in them.

Despite the fact that modelling of water systems by means of ANN has become a common practice abroad, this issue is not sufficiently covered in Russia. A brief overview of modern approaches to the use of ANN in solving water management problems is given. Of the many tasks of this kind, the assessment of the quality of water bodies and monitoring are presented. A detailed analysis of the methodological basis for the creation and application of ANN is beyond the scope of this work.

There are a large number of different ANN. The basic ANN is based on a multi-layer perceptron, i.e. a model of perception and processing of information. ANN usually consists of 3 parallel layers: an input layer, a hidden layer, and an output layer. The input layer contains only input variables. The hidden layer is the most important part of the ANN and contains several neurons that perform certain linear and nonlinear operations on the data. In the output, the result of the transformations is formed [1, 2].

After the structure of the ANN is determined (the number of layers, the number of neurons and the activation function are determined), its "tuning" is carried out by setting the weight coefficients between the input and hidden layer and the offset values (threshold values) of the output neurons. The selection of these parameters in order to increase the reliability of the model is called training or calibration of the INS. To assess the effectiveness (reliability) of the model, and, accordingly, to assess how well certain values of weighting coefficients and threshold values are suitable, statistical confidence indicators are often used – the average square error (the average square of the deviation) and the root of the average square error.
Often, in the works on the use of ANN in solving water management problems, the effectiveness of ANN is compared with other methods of data processing. For example, the application of the support vector machine method is often studied together with the ANN. This method is not an ANN, but it, like an ANN, refers to data analysis methods based on artificial intelligence [3].

The essence of the method is to find the most optimal function for dividing data into classes and constructing the corresponding separating plane. The reference vectors are the points (data) located closest to the dividing plane. Based on them, a separating plane is constructed with a minimum distance (gap) between the reference vectors and the plane, which is the most optimal way to divide the data into classes.

When analysing real data, the optimal data separation function is almost never direct due to large amounts of data, the multidimensional nature of the data space, the need to divide data into many classes, complex correlations between data, etc. Therefore, a hyperplane is often constructed using the support vector method, dividing the data into classes.

2. Results and discussions

2.1. Optimization of monitoring networks

In conditions of high anthropogenic load on water bodies, the creation of wide water quality monitoring networks is an extremely urgent task. Obviously, the most preferable is the widest possible monitoring network, both in terms of the number of control points and in terms of the number of monitored parameters. However, this approach is extremely costly and economically inefficient. This means that the optimal, rather than the maximum, number of control points and controlled water quality indicators should be selected. This approach is valid both for newly created monitoring systems and for optimizing already created systems. The solution of this problem by searching for the most relevant control points and water quality indicators is possible through the use of ANN.

The most preferable is the widest possible monitoring network, both in terms of the number of control points and the number of monitored parameters. However, this approach is economically inefficient. The optimal number is not the maximum number of control points and controlled water quality indicators. This approach is valid both for newly created monitoring systems and for optimizing already created systems. The solution of this problem by searching for the most relevant control points and water quality indicators is possible through the use of ANN.

The paper [4] considers the selection of the most representative water quality indicators, as well as the most important monitoring points, from all the indicators controlled by the monitoring network. The object of the study is the Piabana River in Brazil. The monitoring network consists of 9 stations that monitor 30 water quality indicators. Data for 16 months were analysed. The input INS consist of a matrix of observed concentrations, which uses 13 parameters: aluminium, ammonium, iron, nitrites, sulphates, COD, dissolved oxygen, electrical conductivity, faecal coliform bacteria, total coliform bacteria, temperature, pH, turbidity. The most relevant indicator is faecal coliform bacteria, the least – COD. Monitoring points were also ranked. Special software based on the considered approach is proposed for the operational evaluation of the effectiveness of the monitoring system.

The most common approach to optimizing the amount of data is the principal component analysis (PCA). PCA allows you to reduce the dimension of the data, losing the least amount of information. With regard to water quality monitoring systems, the PCA allows you to determine the most variable water quality indicators and control points where the variability of these indicators is maximum, thereby reducing the potential redundancy of the monitoring system. To date, several such studies have already been conducted, however, they use linear PCA, which can lead to incorrect results, since often the relationships of water data are nonlinear in nature. To solve this problem, it is proposed to use a nonlinear PCA based on auto-associative artificial neural networks together with the general influence technique, which allows us to estimate the level of the general influence of each input variable on each output variable.
In [5], the use of ANN is proposed to optimize the operation of a hierarchical system for monitoring the quality of a water body.

A hierarchical monitoring system is an approach in which the monitoring system has highly informative, expensive sensors that analyse the chemical composition of water, but are able to take a limited number of samples due to the need for analysis using reagents, regular maintenance and calibration. On the other hand, the monitoring system has inexpensive methods for regular monitoring of the state of a water body. The ANN allows you to optimize the use (sampling frequency) of more expensive sensors, based on the data of more affordable sensors.

The object of the study is the River Lee in Ireland. The basis for optimization was the data of the water depth sensor and satellite data on the amount of precipitation. With an increase in precipitation and an increase in the water level in the river, the probability of nutrient substances entering the water, including phosphorus, increases, respectively, during these periods it is necessary to increase the frequency of sampling by a phosphate sensor. Similarly, if there is no or little precipitation and the water level is constant, the probability of nutrients entering the river decreases, which reduces the frequency of sampling by a phosphate sensor. To solve this problem, satellite data on precipitation and data on the depth of the river were processed by the ANN in order to predict the average level of fresh water and optimize the operation of the phosphate sensor. It should be noted that the water level in the studied section of the river is affected not only by the amount of precipitation, but also by the tides (the studied section is located in the area of the river's confluence with the sea) and the dam used for the operation of the hydroelectric power station.

The paper [6] deals with the problem of monitoring remote water bodies, the quality of which is difficult to control by traditional methods. A cost-effective methodology for quality control of remote water bodies based on the processing of remote sensing data of water bodies by neural networks, which does not require the installation of additional monitoring and sampling stations, is proposed.

The object of the study is the Kissimmee River (South Florida, USA), which is subject to serious negative anthropogenic impact in the form of diffuse nitrogen and phosphorus runoff from agricultural areas.

The initial data for the simulation are the results of remote sensing of the river, in various spectra, according to the Landsat system. The most representative from the point of view of eutrophication of the reservoir were considered the content of chlorophyll-a, turbidity and total phosphorus content. It is possible to obtain the values of these indicators by analysing satellite images by analysing the spectrum in the range of 500-600 nm and then-about the amount of chlorophyll in the reservoir. To evaluate the effectiveness of the neural network simulation, data from several monitoring stations were used.

Based on the results of the work, it was shown that the modelling of the content of chlorophyll, phosphorus and modelling of water turbidity by analysing satellite data with a neural network has a high degree of correlation with the data obtained from monitoring stations. Thus, the approach proposed in this paper makes it possible to monitor remote water bodies by analysing satellite data using neural networks.

In [7], a neural network was used to predict the values of dissolved oxygen, total mineralization, total hardness, alkalinity, and turbidity of water based on the values of electrical conductivity, temperature, and pH, followed by modelling other indicators.

In this paper, the effectiveness of nonlinear (neural network) and linear (mathematical modelling based on the polynomial approximation with the least squares method) modelling was compared. The data sources were 30 water quality monitoring stations located in the Bay area of Michigan.

According to the results of the work, the neural network showed a high degree of data reliability, and surpassed the indicators of linear modelling. In addition, the authors note that regardless of the type of modelling, the quality and accuracy of the initial data is of critical importance.

The efficiency was compared using the ANN and the polynomial approximation. The neural network showed a higher degree of data reliability.
2.2. Optimization of water treatment and purification processes

Cleaning of polluted waters is one of the most important measures in the water management system. At the same time, like any other activity, water treatment should be cost-effective. This means that a strictly optimal amount of reagents must be used in the purification process, and the final water quality must meet regulatory requirements. To optimize the water treatment process, it is possible to use modelling of the behaviour of the water environment, treatment reagents and pollutants. Given the variety of hidden nonlinear connections, INS successfully copes with this task.

In [8], the problem of optimizing the amount of chlorine used in the water treatment process is considered. The use of an insufficient amount of chlorine will not ensure proper water quality, and the use of an excessive amount of chlorine will lead to an increase in its concentration in the water that has been treated, which will reduce its quality. Thus, it is necessary to ensure the use of a strictly defined optimal amount of chlorine in the water treatment process.

A model was proposed to predict the content of residual chlorine. The following data modelling and analysis tools were considered:

1. Artificial neural networks, including:
2. The method of support vectors;
3. Classification and regression tree.

Several statistical indicators were used to assess the quality of the model:

According to the results of the work, it was found that the model based on a multilayer perceptron neural network has the greatest reliability.

The paper [9] presents a simulation of water quality indicators that have been treated at a water treatment plant in order to optimize the operation of the station. Water treatment plants are complex dynamic systems, the condition of which must be monitored on a regular basis for timely adjustment of technological processes and optimization of their functioning. However, the traditional approach to assessing the effectiveness of technological processes of water treatment often requires significant time costs. For example, the determination of the BOD of water requires 5 days of incubation of the sample. To solve such problems, modelling of the functioning of water treatment plants is used. In this paper, we consider the use of neural networks for modelling BOD, COD and total nitrogen in treated waters. The initial data for modelling included the following indicators of the quality of water entering for treatment: pH, electrical conductivity, BOD, COD, total nitrogen content.

To evaluate the effectiveness of the model and the method of constructing ensembles, the determination coefficient and the root of the mean square error were used.

According to the results of the work, it was found that the constructed ANN model based on the fuzzy inference system has the greatest efficiency.

In [10], the authors created a hybrid statistical model based on a neural network and a genetic algorithm (an evolutionary algorithm for finding optimal modelling parameters), which allows predicting the average monthly productivity of water treatment stations of a drinking water supply system based on water quality indicators and production indicators of the station (input parameters of the model).

It should be noted that in terms of the functioning of water treatment plants, often only individual technological processes are modelled, while in this paper the functioning of the water treatment plant as a whole is modelled. The data were obtained from 45 water treatment stations of the drinking water supply system of China. The genetic algorithm was used to more accurately determine the weight and threshold values of the neural network, i.e. to optimize its operation. Based on the results of the work, the authors concluded that the data generated by the developed hybrid model is of high quality. The data quality increases in proportion to the amount of initial (training) data, which indicates the further potential for the development of the proposed approach. The created model is a promising tool for predicting the performance of water treatment plants based on the assessment of the technological indicators of the station and the indicators of the quality of water supplied for treatment.
2.3. Modelling of the state of aquatic ecosystems

The modelling of the most representative indicators of water quality, which indicate the state of the quality of the water body as a whole, is relevant.

The paper [11] presents the simulation of the amount of dissolved oxygen in reservoirs by the INS method. The object of the study is the Feitsu Reservoir, Taiwan. The following parameters were used as input data: water temperature, pH, electrical conductivity, turbidity, suspended particle concentration, total hardness, alkalinity, and ammonium concentration.

The paper [12] presents a simulation based on various types of neural networks of the content of chlorophyll in water bodies as a key indicator of eutrophication of water bodies. Baseline data were obtained for 1,000 water bodies (lakes, ponds, reservoirs) in the United States during the National Lake Assessment Program for 2007 and 2012. To predict the chlorophyll content, the following water quality indicators were used: transparency, turbidity, total nitrogen content, and total phosphorus content. Reservoirs of natural and artificial origin were considered.

According to the results of the work, it was found that the quality indicators of natural reservoirs are more representative. Neural networks based on a multilayer perceptron were the most effective for modelling.

2.4. Modelling of diffuse pollution

Diffuse pollution contributes significantly to the total amount of water pollution. At the same time, the control of diffuse pollution is a difficult task due to the spatial distribution of pollution sources. ANN is a promising tool for modelling diffuse pollution.

In [13], the authors compared the efficiency of forecasting the total nitrogen and phosphorus content in the Changle River, Southeast China, which is characterized by a high level of diffuse pollution. The data were obtained on a monitoring network consisting of 12 sampling stations. The following values were used as input parameters of the model: water temperature, flow velocity, precipitation, dissolved oxygen concentration, total nitrogen and phosphorus content upstream of the river.

In [14], the sources of diffuse pollution of the Liao River, China, were evaluated. Two models were combined: a neural network model of precipitation as the main mechanism of diffuse pollution and a SWAT model (soil and water assessment tool) for modelling the behaviour of the "soil-water" system and estimating the volume of diffuse pollution entering the river. The training sample for predicting the amount of precipitation by the neural network included a 10-year period of operation of 8 precipitation monitoring posts. As a result, the areas with the highest amount of precipitation and the greatest agricultural activity were identified, which are the main sources of diffuse pollution of the river.

3. Conclusions

The analysis of modern approaches to solving some water management problems has shown that:

– Neural networks can solve a wide range of forecasting and optimization problems in the presence of a large amount of data.

– The main advantage of neural networks is their ability to identify and process nonlinear relationships between variables.

– Models based on neural networks can become an important tool for water management.

Acknowledgment

The study was carried out within the framework of the state task AAAA-A18-118022090106-2 and funded by RFBR, project number 19-05-50082.

References

[1] Zhu S, Heddam S, Nyarko E K et al. Environ Sci Pollut Res 2019 Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and
artificial neural networks models Environmental Science and Pollution Research 26(1) pp 402–420

[2] Giancarlo Zaccone, Getting Started with Tensor Flow, Packt Publishing, 2016

[3] Ehteram M, Afan H A, Dianatikhah M, Ahmed A N, Ming Fai C, Hossain M S, Allawi M F and Elshafie A 2019 Assessing the Predictability of an Improved ANFIS Model for Monthly Streamflow Using Lagged Climate Indices as Predictors Water 11 p 1130

[4] Villas-Boas M D, Olivera F and de Azevedo J P S 2017 Environ Monit Assess 189 p 439

[5] O’Connor E, Smeaton A F, O’Connor N E, Regan F 2012 A neural network approach to smarter sensor networks for water quality monitoring Sensors (Basel) 12(4) pp 4605–4632

[6] Chebud Y, Naja G M, Rivero R G and Melesse A M 2012 Water quality monitoring using remote sensing and an artificial neural network Water, Air, & Soil Pollution 223(8) pp 4875–4887

[7] Salari M, Salami Shahid E, Afzali S H, Ehteshami M, Conti G O, Derakhshan Z and Sheibani S N 2018 Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water Food and Chemical Toxicology 118 pp 212–219

[8] Zounemat-Kermani M, Ramezani-Charmahineh A, Adamowski J and Kisi O 2018 Investigating the management performance of disinfection analysis of water distribution networks using data mining approaches Environmental monitoring and assessment 190(7) pp 1-15

[9] Nourani V, Elkiran G and Abba S I 2018 Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach Water Science and Technology 78(10) pp 2064-2076

[10] Zhang Y, Gao X, Smith K, Inial G, Liu S, Conil L B and Pan B 2019 Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network Water Research 164 p 114888

[11] Chen W B and Liu W C 2014 Environ Monit Assess 186 p 1203

[12] Luo W, Zhu S, Wu S et al. 2019 Comparing artificial intelligence techniques for chlorophyll-a prediction in US lakes Environmental Science and Pollution Research 26(29) pp 30524–30532

[13] Liu M and Lu J 2014 Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river Environmental Science and Pollution Research 21(18) pp 11036–11053

[14] Wang Y, Bian J M, Wang S N and Nie S Y 2016 Predicting precipitation on nonpoint source pollutant exports in the source area of the Liao River, China Water Science and Technology 74(4) pp 876–887