# Water Quality Evaluation Model Based on Hybrid PSO-BP Neural Network

**Yan Wu[1], BingXiang Liu[1], Xing Xu*[,1,2], Na Hu[1], Kezhong Tang[1], Mengshan Li[1]**
[1]Information engineering college,Jingdezhen ceramic institute, Jingdezhen, JiangXi, 333403
[2]State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China;
*Corresponding author, e-mail: whuxx84@aliyun.com

***Abstract***

*A hybrid neural network algorithm, aims at evaluating water quality, based on particle swarm optimization (PSO) algorithm, which has a keen ability in global search and back propagation (BP) algorithm that has a strong ability in local search. Heuristics has been proposed to optimize the number of neurons in the hidden layer. The comparison with the traditional BP NN shows the advantage of the proposed method with high precision and good correlation. The values of average absolute deviation (AAD), standard deviation error (SDE) and squared correlation coefficient (R2) are 0.0072, 0.0208 and 0.98845, respectively. The results show that the hybrid PSO-BP NN has a good predical ability of evaluating water quality; it is a practical and efficacious method to evaluate water quality.*

*Keywords*: hybrid PSO-BP model, water quality, PSO algorithm, BP neural network, heuristics

## 1. Introduction

Water quality evaluation is not only an important step in System of Water-quality, but also a very important and essential concept in the assessment on comprehensive environmental quality. In recent years, the water quality evaluation has attracted more and more attention. In order to use and protect water resources effectively, some researchers work on the pollution degree and the development trend of lake and river. The traditional methods of water quality assessment include: Single Factor Water Quality Identification Index [1], the Principle component analysis [2] and the comprehensive water quality identification index [3]. Single Factor Water Quality Identification Index which employs the worst single factor of water quality evaluates comprehensive water quality grades. This assay method is simple and shows the relationship between Water Quality and evaluation standard. But it doesn't show the comprehensive evaluation result, whose disadvantages lie in bad accuracy. The Principle component analysis which uses comprehensive evaluation factors establishes a comprehensive evaluation model for water quality. The result is ideal. But if there are too many factors involved in the model, the contribution rate of principal components will be reduced. It is difficult to achieve a better prediction effect. The water quality can be influenced by multiple factors, so there is a complicated non-linear relation between evaluation Index and water quality standard. The traditional analysis methods are difficult to obtain the good evaluation results of complicated nonlinear problem. With the rapid development and wide use of information technology and etc, the traditional analysis methods should be replaced by intelligent optimization algorithms.

Research on Artificial Neural Networks (ANN) for water quality assessment and simulation is also a research hotspot in recent years. The method is simple in structure, but can most solve linear or non-linear practical problems, and has been applied extensively in water quality assessment model. Cao Yanlong [4] has established the model of BP neural network (BPNN) that focuses on real problems, but this kind of method contains its flaws and shortages, such as the poor convergence, and the fall in local optima. Dong Manling [5], Li Jiake [6] presented a radial basis function neural networks for water quality assessment. Hua Zulin [7] proposed the LM-BP neural networks model for water quality assessment. Chen Shouyu [8] proposed the fuzzy Artificial Neural Network (FANN) model for water quality assessment. To sum up the above arguments, modified BP algorithms improve the constringency speed and forecasting precision. But there is still room for improvement.

In the application of ANN, the determination of neural network structure, parameters and bias is the most important and crucial factor because the training process of ANN could be considered as a classical optimization problem [9]. Recently, researchers discovered that many intelligent algorithms, such as genetic algorithm (GA) [10, 11], simulated annealing algorithm (SA) [12], fuzzy theory [13], wavelet analysis [14], particle swarm optimization algorithm (PSO) [15] and chaos theory [16], can be used for this determination. Therefore, the ANN combined with intelligent optimization algorithms, and also referred as hybrid neural network, has become a very active subject in recent years [17].

At present, PSO has become the hot research of AI field .The PSO algorithm is a global and advanced algorithm with a strong ability to search the global optimistic result. However, the PSO algorithm has a disadvantage that it is easy to sink into local optima [18]. Fortunately, on the contrary, the BP algorithm has a dazzling ability to find local optimistic result [19]. Therefore, we have focused on the hybrid neural network based on PSO algorithm and BP algorithm (PSO-BP) in order to form a hybrid learning algorithm for training an ANN model. The superiority of the hybrid PSO-BP algorithm is that the combinability of the PSO algorithm's keen ability in global search and the BP algorithm's strong ability in local search. In this work, an ANN model based on hybrid PSO-BP algorithm has been established for the water quality in a wide rang of pollution indicators. In order to evaluate the performance of the proposed hybrid neuron network, the predicted results in this work are compared with BP NN.

## 2. The Relative Theory
### 2.1. Back Propagation Neural Network and Particle Swarm Optimization Algorithm
Research shows that the BP NN can simulate almost all the linear and nonlinear problems and can predict the enormously complicated and multivariable processes with high precision depending on the network been trained completely [20]. The BP NN learning process which consists of input signals feed forward and error signals back propagation. Firstly, the input signals transmit to the output layer over the hidden layer, the output layer exports results which respond to the predicted values, in this process, the connection weights remain unchanged [19, 21]. Secondly, the error signals, the differences between the actual values and the predicted values, turn back from output layer, during this process, the connection weights are constantly adjusted by means of error signals with the purpose of getting error minimum.

The PSO algorithm works by initializing a flock of birds randomly in the search space. Every bird is called as a particle represented as a potential solution, flies through the search space with a certain velocity following the current optimum particles, and finds the best global position after some iteration. At each iteration, each particle can adjust its velocity and position vector based on its momentum and the influence of its best position (Plbest) and the best position of its neighbors (Pgbest) [9, 20].

In PSO algorithm, assuming in an n-dimensional search space, the total number of particles is m, vector $X = (x_1, x_2, \cdots, x_m)^T$, the position of the ith particle can be expressed as vector $x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,n})^T$, the velocity of the ith particle can be expressed as vector $v_i = (v_{i,1}, v_{i,2}, \cdots, v_{i,n})^T$, the best position of the ith particle $p_i = (p_{i,1}, p_{i,2}, \cdots, p_{i,n})^T$, and the best position of the neighbor particles $p_g = (p_{g,1}, p_{g,2}, \cdots, p_{g,n})^T$. The position and velocity are updated by using the equations.

$$v_{i,d}^{k+1} = v_{i,d}^{k} + c_1 rand()(p_{i,d}^{k} - x_{i,d}^{k}) + c_2 rand()(p_{g,d}^{k} - x_{i,d}^{k}) \tag{1}$$

$$x_{i,d}^{k+1} = x_{i,d}^{k} + v_{i,d}^{k+1} \tag{2}$$

Where i=1,…,m; $x_{i,d}^{k}$ and $v_{i,d}^{k}$ are the position and velocity of ith particle at d-dimensional and the kth iteration; c1 and c2 are the acceleration constants with positive values,

also called training factor; $p_{i,d}^{k}$ is the best position of ith particle in d-dimensional while $p_{g,d}^{k}$ is the best global position.

## 2.2. Hybrid PSO-BP NN Algorithm

The PSO-BP algorithm is an optimization algorithm based on the PSO algorithm and the BP algorithm. The search process of the PSO-BP algorithm starts from initializing a group of particles randomly. All particles in the search space are updated according to the Equation (1) and (2).

At every iteration, a new group of particles are generated for searching the best global position in the solution space. Then the local optimistic BP algorithm is used to search around the global optimum that the PSO algorithm has searched. In this way, this hybrid PSO-BP algorithm may find an optimum more efficiently and quickly [20]. Hybrid PSO-BP Neural Network flow chart, as shown in Figure 1.
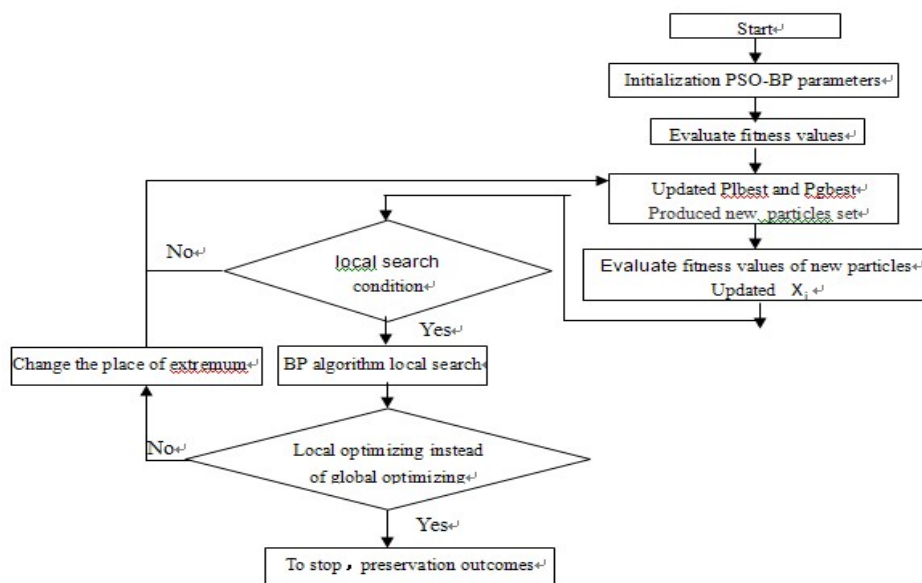


Figure 1. Flow Chart of Hybrid PSO-BP NN

## 3. Experiment and Conclusion

The predictability of the model was evaluated by calculating average absolute deviation (AAD), standard deviation error (SDE) and squared correlation coefficient ($R^2$). The AAD and SDE are defined as:

$$AAD = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\Pr e(i) - Exp(i)\right|}{Exp(i)}$$

$$SDE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x\right)^2}$$

Where N is the number of data points; $Pre(i)$ is the predicted value of model and $Exp(i)$ is the experimental data; the $x$ is the average of the N data points.

## 3.1. Model Structure

As mentioned before, the data points preparative beforehand can be excavated in the input and output parameters. The hybrid PSO-BP NN model with three layers, one input layer with four nodes represented Dissolved Oxygen, volatile phenol, permanganate index and

ammonia nitrogen, one hidden layer, and one output layer with one node which represented the water quality grade was designed in this work.

In allusion to the hidden layer, the number of neurons in the hidden layer wholly depends on the nature of the specific problem. Therefore, it is necessary to optimize this parameter with the purpose of searching the best value of the number of neurons. In this paper, heuristics have been proposed to optimize the number of neurons in the hidden layer. Utilizing the previously prepared procedure, 11 hybrid PSO-BP NN modules were generated assuming the number of neurons in the hidden layer from five to fifteen. Table 1 shows the AAD, SDE, R2 and the best fitness calculated for different network configurations, differing with respect to the number of neurons in their hidden layer. As shown in the Table 1.

Table 1. Topology Studies for Optimal ANN Configuration

| Hidden Neuron | AAD | Best Fitness | $R^2$ | SDE |
|---|---|---|---|---|
| 5 | 0.0194 | 1.02E-03 | 0.9613 | 0.0318 |
| 6 | 0.0126 | 1.01E-03 | 0.9711 | 0.0417 |
| 7 | 0.0107 | 9.89E-04 | 0.9726 | 0.0412 |
| 8 | 0.0092 | 6.44E-05 | 0.9817 | 0.0309 |
| 9 | 0.0072 | 1.00E-06 | 0.9884 | 0.0208 |
| 10 | 0.0093 | 9.79E-05 | 0.9714 | 0.0210 |
| 11 | 0.0106 | 9.81E-05 | 0.9707 | 0.0223 |
| 12 | 0.0164 | 9.74E-05 | 0.9748 | 0.0245 |
| 13 | 0.0177 | 1.14E-04 | 0.9715 | 0.0425 |
| 14 | 0.0165 | 1.28E-05 | 0.9625 | 0.0459 |
| 15 | 0.0272 | 1.55E-04 | 0.9673 | 0.0463 |

Generally, the network with the least error and the most suitable correlation coefficient has been chosen as the optimal network configuration [9, 25]. In the present work, it is obvious that the network with the lowest AAD and SDE, and the highest R2 has been selected. According to the bold row of Table 1, the hidden layer with nine neurons was selected for the hybrid PSO-BP NN.

It is fortunate that the PSO algorithm has few parameters to be adjusted in an actual application. However, the only several parameters have an important effect on performance and convergence of PSO algorithm. For the present, the parameter setting of PSO depends on experience to a great extent [9, 20]. The main parameters calibrated in the model are population size (N=50), iteration number (itmax=1000), the least error (minerror=1.00E-06), the maximum weight (wmax=0.9), the minimum weight (wmin=0.3)and learning factor (c1=2.8，c2=1.3).

### 3.2. Experimental Data

According to the surface water environment quality standard (GB3838-2002) of China, in this paper, we will generate the training samples which include 218 sets of data. The training samples are made by unitary processing according to maximum or minimum.

$$y(k) = (x(k) - \min(x(n))) / (\max(x(n)) - \min(x(n))), \quad k = 1, 2, \mathrm{L} \ N$$

Where min(x(n)) is minimum of sample data x(n), max(x(n)) is maximum of sample data x(n).In this paper, carrying on the test simulation to 10 test sample data that the key pollution indicators are Dissolved Oxygen, volatile phenol, permanganate index and ammonia nitrogen evaluate theforecasting model. As shown in the Table 2.

### 3.3. Results and Discussion

In this study, a three-layer feed forward artificial neural network (ANN) model based on hybrid algorithm combining PSO with BP was proposed to predict the water quality. For this

purpose, we have built a 4-9-1 hybrid PSO-BP NN model, the input layer with four nodes represented Dissolved Oxygen, volatile phenol, permanganate index and ammonia nitrogen; nine neurons in the hidden layer and one node represented the water quality grade in the output layer.

Table 2 Measurement Data for Water Quality Evaluation

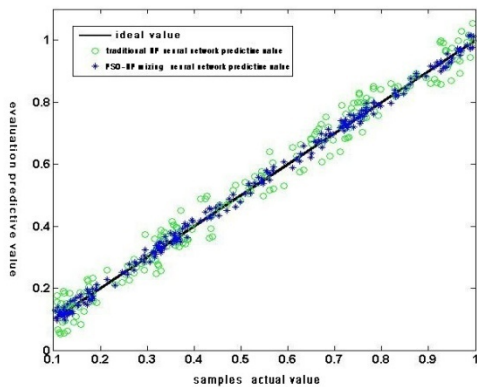| Simple | DO | Volatile phenol | CODmn | NH3-N | Oupput |
|--------|-----|-----------------|-------|-------|--------|
| 1 | 7.9 | 0.001 | 1.7 | 0.21 | 0.3 |
| 2 | 8.0 | 0.003 | 3.4 | 0.15 | 0.7 |
| 3 | 7.2 | 0.002 | 3.0 | 0.17 | 0.3 |
| 4 | 7.6 | 0.001 | 2.4 | 0.38 | 0.3 |
| 5 | 3.6 | 0.034 | 19.5 | 11.2 | 0.7 |
| 6 | 5.1 | 0.006 | 64.4 | 4.78 | 0.9 |
| 7 | 3.1 | 0.019 | 16.4 | 2.91 | 0.7 |
| 8 | 4.1 | 0.004 | 6.6 | 3.10 | 0.9 |
| 9 | 5.4 | 0.001 | 2.9 | 0.30 | 0.3 |
| 10 | 6.2 | 0.001 | 3.2 | 0.49 | 0.3 |



Figure 2. Comparison between Measurement and Prediction in Training Set
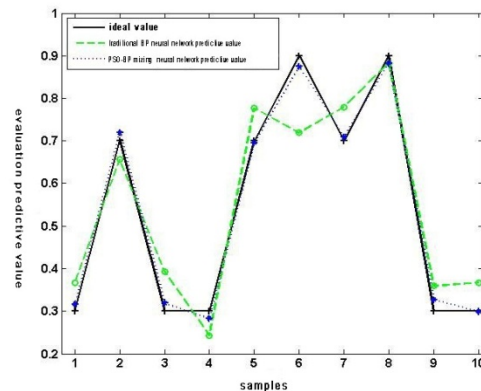
Figure 3. Comparison between Measurement and Prediction in Testing Set

The original data shown in Table 2, which consist of four key pollution indicators, were the experimental data extracted from literature. In Figure 2 and Figure 3, the prediction of water quality grade by hybrid PSO-BP NN model and BP NN model are plotted against the experimental data for the training and testing sets. In these figures, lines show the ideal modeling that the prediction values equal to the experimental data. In Figure 2, the asterisk and circle are the correlations between the actual sample values and predictable values in training set by PSO-BP NN model and BP NN model. In Figure 3, the broken line and dotted line are the correlations between the actual sample value and predictable values in testing set by PSO-BP NN model and BP NN model.

As shown in these figures, the outputs of the hybrid PSO-BP NN model show a fair agreement with the experiment, no matter what the training set or the testing set is. The simulating performance of the hybrid PSO-BP NN model was evaluated by comparing with the traditional BP NN model established beforehand based on calculating AAD, SDE and R2. It is certain that a fair comparison is achieved when the two models are based on the same data, so the testing sets were employed in this comparison. Table3 shows the AAD, R2 and SDE for the two models based on testing sets. The statistical results of the two models are compared in Table3. It reveals that, in the uniform data set, the hybrid PSO-BP NN model is more accurate than traditional BP NN model. The major reason for the superiority of the hybrid PSO-BP NN

model is the combinability of the particle swarm optimization algorithm's keen ability in global search and the back-propagation algorithm's strong ability in local search.

Table 3. Comparison Data between Hybrid PSO-BP NN and BP NN

| Sample | BP NN | | | PSO-BP NN | | |
|---|---|---|---|---|---|---|
| | AAD | $R^2$ | SDE | AAD | $R^2$ | SDE |
| training set | 0.0206 | 0.9676 | 0.0717 | 0.0071 | 0.9886 | 0.0207 |
| Test set | 0.0233 | 0.9587 | 0.0808 | 0.0073 | 0.9883 | 0.0209 |
| Average | 0.02195 | 0.96315 | 0.07625 | 0.0072 | 0.98845 | 0.0208 |

## 4. Conclusion

Recently, predicting the water quality by ANN model becomes the most vital subject in studying on water quality model. In this paper, we have proposed a hybrid neural network, which based on the PSO algorithm and the BP algorithm to predict the water quality. The potent ability of hybrid PSO-BP NN in global and local search could get the better predictable results compared to the traditional BP NN. The performance of hybrid PSO-BP NN was evaluated based on calculating the AAD, R2 and SDE. The results indicate that the hybrid PSO-BP NN model is a reliable, accurate and applied model for the water quality evaluation, and it is also a practicable method for the analysis and design of the water quality model.

This study indicates that constructing a hybrid PSO-BP NN model for predicting the water quality has a good prospect for application. In the future research works, we will follow up on this subject all the time and focus on how to apply this hybrid neuron network to solve more realistic problems.

## References

[1] Wang M, Zhou W, Hu C. Status of nitrogen and phosphorus in waters of Lake Poyang Basin. *Journal of Lake Sciences*, 2008; 20(3): 334-338.
[2] Ou J, Ye J, Cheng J. Application of main component analysis method in comprehensive evaluation for water conservancy modernization in Jiangsu Province. *Yangtze River*. 2010; 41(2): 97-100.
[3] Wang H, Peng J, Liang K. Application of comprehensive water quality labeling index method in assessment of reservoir water quality. *Yangtze River*. 2007; 38(1): 92-94.
[4] Cao Y, Wang X, Zhou Z. Water quality assessment method of Wei fiver based on BP neural network. *Computer Engineering and Design*. 2008; 29(22): 5910-5916.
[5] Dong M, Huang S. Application of BF Network to Evaluate Water Quality. *Environmental Science & Technology*, 2003; 26(01): 23-25.
[6] Li Jiake, Zhou Xiaode, Li Yajiao, etc. The Application of RBF Neural Network in Water Quality Evaluation. *Journal of Irrigation and Drainage*. 2003; 22(6): 70-73.
[7] Hua Z, Qian W, Gu L. Application of improved LM-BP neural network in water quality evaluation. *Water Resources Protection*. 2008; 24(4): 22-25.
[8] Chen S, Li Y. Water quality evaluation based on fuzzy artificial neural network. *Advances in Water Science*. 2005; 16(1): 88-91.
[9] Mohammad A A. Neural network based unified particle swarm optimization for prediction of asphaltene precipitation. *Fluid Phase Equilibria*, 2012; 314 (2012): 46-51.

[10] Mehdizadeh B, Movagharnejad K. A comparative study between LS-SVM method and semi empirical equations for modeling the solubility of different solutes in supercritical carbon dioxide. *Chemical Engineering Research and Design*. 2011; 89(11): 2420-2427.

[11] Valderrama JO, Zavaleta J. Thermodynamic consistency test for high pressure gas–solid solubility data of binary mixtures using genetic algorithms. *J. Supercrit. Fluid*, 2006; 39: 20-29.

[12] Niaki S, Nasaji S. A hybrid method of artificial neural network and simulated annealing in monitoring autocorrelated multi-attribute processes. *Int J Adv Manuf Technol*. 2011; 56:777–788.

[13] Khajeh A, Modarress H. Prediction of solubility of gases in polystyrene by adaptive neuro-fuzzy inference system and radial basis function neural network. *Expert Systems with Applications*. 2010; 37 (2010): 3070-3074.

[14] Khayamian T, Esteki M. Prediction of solubility for polycyclic aromatic hydrocarbons in supercritical carbon dioxide using wavelet neural networks in quantitative structure property relationship. *Journal of Supercritical Fluids*. 2004; 32(1): 73-78.

[15] Lazzus JA. Estimation of solid vapor pressures of pure compounds at different temperatures using a multilayer network with particle swarm algorithm. *Fluid Phase Equilibria*. 2010; 289(2): 176-184.

[16] Sun YB, Babovic V, Chan ES. Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *Journal of Hydrology*. 2010; 395(1-2): 109-116.

[17] Jesuthanam C P, Kumanan S, Asokan P. surface roughness prediction using hybrid neural networks. *Machining Science and Technology*, 2007; 11(2) 271-286.

[18] Yan X, Wu Q, Liu H. Orthogonal Particle Swarm Optimization Algorithm and Its Application in Circuit Design. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(6): 2926-2932.

[19] Maier HR, Dandy GC. The effect of internal parameters and geometry on the performance of back-propagation neural networks: an empirical study. *Environmental Modelling and Software*. 1998; 13(2): 193-209.

[20] Lashkarbolooki M, Vaferi B, Rahimpour MR. Comparison the capability of artificial neural network (ANN) and EOS for prediction of solid solubilities in supercritical carbon dioxide. *Fluid Phase Equilibria*. 2011; 308(2011): 35-43.

[21] Bakhbakhi Y. Neural network modeling of ternary solubilities of 2-naphthol in supercritical CO2: A comparative study. *Mathematical and Computer Modelling*. 2012; 55(7-8): 1932-1941.