Acta Technica 62 No. 2A/2017, 131–142

Study on prediction and evaluation of Chaohu Lake water quality with MPSO-FSMV

CHI ZHANG[1]

Abstract. In consideration of the low prediction accuracy, poor applicability and other weaknesses in traditional water prediction and evaluation models, to realize the high-accuracy water prediction and evaluation, the modified inertia weight was introduced into the particle swarm optimization (PSO) based on fuzzy support vector machine (FSVM) and a MPSO-based optimized FSVM water prediction and evaluation model was proposed. The high-accuracy water prediction and evaluation can be achieved in the condition of optimal parameters by optimizing Gamma and b, FSVM parameters, with MPSO. With the monitoring data of Chaohu Lake water quality from 2010 to 2015 as the object of study, we realized the prediction and evaluation on water quality of Chaohu Lake based on studying the variation trend and time-space relationship of water quality. It is found through comparison of 4 types of water quality evaluation and prediction methods including MPSO-FSVM , PSO-FSVM , FSVM and SVM that the prediction accuracy of MPSO-FSVM reaches up to 95.38 %, which is higher than the prediction accuracies of PSO-FSVM , FSVM and SVM. Therefore, it indicates that MPSO-FSVM has a higher precision and adaptability in water quality evaluation and prediction; the validity and reliability of MPSO-FSVM algorithm was verified, so it can be popularized to other fields of scientific researches and engineering applications.

Key words. PSO, FSVM, water quality evaluation, variation trend, neural network, grey prediction.

1. Introduction

Water quality evaluation and prediction, which are significant research topics and important contents of the foundation of modern environmental science, aim to correctly reflect the quality and pollution of water environment via water quality evaluation, predict the development tendency of water environment quality and thus provide a scientific decision-making basis and method for management, protection and governance of water environment. As the fifth biggest freshwater lake of China, Chaohu Lake, which is located in the central hilly land between the Yangtze River and Huaihe River in Anhui Province, plays an extremely important role in the

[1]Shandong University of Science and Technology, Shandong, 266590, China

economic development of Anhui, so it is of great significance to predict and evaluate the water quality condition of Chaohu Lake.

At present, the methods for water quality prediction mainly consist of the grey prediction method [1], artificial neural network [2] and SVM [3] etc. Although a good result can be obtained by using the grey prediction method for water quality prediction, a high grey level of data will lead to low prediction accuracy; meanwhile, this method is not suitable for long-term prediction of water quality [4–5]. In order to enhance the accuracy in water quality prediction, the literature [6] and literature [7] proposes water quality prediction models respectively based on weighted array, exponential smoothing and GM(1,1) model combination; they all can improve the prediction accuracy to some degree but this method fails to fundamentally solve the problem of prediction error. To enhance the accuracy in water quality prediction, literature [8] puts forward a water evaluation model based on grey metabolism GM(1,1) model. As shown in the empirical result, the prediction accuracy with the method is superior to the traditional GM(1,1) model.

Owing to its accurate nonlinear mapping ability and generalization ability, the artificial neural network has been widely used in scientific research and engineering application. Based on the mapping relation between organic maters in surface river water and main influence factors, literature [9] proposed a water quality prediction model based on BP neural network. The experimental result shows that the method has a high accuracy and generalization ability. Since the small sample data based on BP neural network has low prediction accuracy and is easy to lead local optimum, literature [10] proposed a water quality prediction model based on an improved BP neural network, which improves the prediction accuracy. As for the problem of local optimum of BP neural network, literatures [11–12] optimized the weight, threshold value and structure of BP neural network with the genetic algorithm and established the water quality prediction model of BP neural network improved by the genetic algorithm. It is thus found through comparison that, after optimization with the genetic algorithm, the prediction accuracy and algorithm stability are better than those of the standard BP neural network.

As for the small sample, nonlinear and high-dimension data, the SVM has its unique advantages, so many scholars introduced it into water quality evaluation and made great achievements. However, selection of kernel function and parameter of SVM directly influences the prediction result [13–15]. Since the parameter selection of single SVM prediction model has the disadvantages of low efficiency and dependence on experience, literature extracted the data feature information of water quality with the wavelet analysis and proposed a water quality prediction model based on wavelet transformation and SVRM (Support Vector Regression Machine). Literature brought forward a water quality parameter prediction model based on improved WSVM (weighted support vector machine); the prediction accuracy of this improved method is superior to the unimproved SVM and BP algorithm. As for the small sample and jumping time series data of water quality, literature proposed a water prediction model based on ELPM data pre-processing and LSSVM parameters optimized by simulated annealing algorithm.

Because there are disadvantages such as low prediction accuracy and poor adaptability in the traditional water quality prediction and evaluation model, this paper introduced the modified inertia weight in the PSO based on FSVM and brought forward a water prediction and evaluation model based on FSVM improved by MPSO.

2. FSVM

In 2002, Lin et al. came up with the FSVM algorithm. When the FSVM is used for classification, compared with the training samples of conventional SVM, apart from the features and type identification of sample, the fuzzy membership degree is added to each sample trained with FSVM to reduce the influence of noise point. Let the training sample set be $(x_1, y_1, \mu(x_1))$, $(x_n, y_n, \mu(x_n))$, $x_i \in R^N$, $y_i \in \{-1, 1\}$, $0 < \mu(x_i) \leq 1$. Suppose that $z = \phi(x)$ is the mapping relation by mapping the training sample from the original space RN to the high-dimensional feature space Z. The fuzzy membership degree $\mu(x_i)$ shows the degree of reliability of the sample in a certain type; i signifies the error item in classification in the objective functions of SVM; then, $\mu(x_i)\xi_i$ is the weighted error item. From literature, it is concluded that the optimal classification plane is the optimal solution to the objective function in the formula

$$\begin{cases} \Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}[\mu(x_i)\xi_i], \\ \text{s.t. } y_i\left[\left(w^{\mathrm{T}} z_i\right) + b\right] - 1 + \xi_i \geq 0, \ i = 1, \cdots, n, \\ \xi_i \geq 0, i = 1, \ \cdots, n \end{cases} . \tag{1}$$

In formula (1), the penalty factor C is a constant; w signifies the weighted coefficient of linear classification function y_i. The discrimination function formula of corresponding optimal face is

$$f(x) = \operatorname{sgn}\left(\sum_{x_i \in SV} w_i K(x_i, x) + b\right). \tag{2}$$

In formula (2), $K(x_i, x)$ signifies the kernel function. The frequently-used Gaussian kernel function is adopted in this paper and now the formula is

$$K(x, z) = \exp\left(-\frac{\mathit{\Gamma} \cdot \|x - z\|^2}{2}\right), \tag{3}$$

where $\mathit{\Gamma}$ refers to the Gaussian distribution width.

Selection of a proper membership degree in a given problem will directly affect the effect of classification. At present, there are various methods to determine the membership degree, such as linear function, quadratic function, heuristic method and noise distribution. This paper introduces a simple and effective infinitely continuous and differentiable membership function.

If $S^+ = \{x_i | y_i = 1\}$ signifies the positive sample set and $S^- = \{x_i | y_i = 0$

signifies the negative sample set, the sample novelty membership function is shown below:

$$\mu_i(z)\begin{cases} b\sigma + b\exp\left(\frac{-1}{r^2-\|z-c\|^2}\right) \|z-c\| < r, \\ \\ b\sigma\,(\text{Other})\,. \end{cases} \tag{4}$$

In the formula above, $b = \left[\sigma + \exp\left[\frac{-1}{r^2}\right]\right]^{-1}$ is a sufficiently small positive number; the function is a infinitely continuous and differentiable function, $||z-c||$ signifies the distance between the two points, c is the center of a certain type, and r is the radius of the smallest hypersphere containing the sample set. If the distance from the central point is shorter, the corresponding value is bigger. The value corresponding to the point when $||z-c|| = r$ is the smallest one. As a general rule, there is a long distance between the noise point and the center of such points or $||z-c|| = r$. As for the noise point, the value (fuzzy relation) calculated according to formula (4) is very small, so there is a small influence on correct classification. Thus, the influence of noise on the classification result can be reduced and the FSVM classification accuracy can also be enhanced.

3. MPSO

3.1. PSO

PSO is a swarm intelligence algorithm proposed with the inspiration of bird flock's foraging behavior. In the algorithm, the particle signifies the solution vector and the quality of particle is determined according to the fitness function size. On this basis, continuous renewal of particle position and speed can be realized, and the global optimal searching and optimization can be attained. $X_i = (x_{i1}, x_{i2}, \cdots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, \cdots, v_{id})$ signify the particle positions and speeds, respectively. The updating strategy for particle position and speed is shown in the formulas

$$V_{id}^{(t+1)} = \omega \cdot V_{id}^{(t)} + c_1 r_1 \left(p_{id}^{(t)} - X_{id}^{(t)}\right) + c_2 r_2 \left(p_{gd}^{(t)} - X_{id}^{(t)}\right)\,, \tag{5}$$

$$X_{id}^{(t+1)} = X_{id}^{(t)} + V_{id}^{(t+1)}\,, \tag{6}$$

and

$$\omega{=}\omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \cdot t\,. \tag{7}$$

In the above formulas (5)–(7), $p_{id}^{(t)}$ and $p_{gd}^{(t)}$, respectively, denote the individual optimal solution and global optimal solution of particle at tth moment of iterations, $r_1, r_2 \in (0,1)$ denote the random numbers, c_1, c_2 denote the learning factors; $t_{\max}$ and t, respectively, denote the maximum iterations and current iterations, w signifies the inertia weight, and $w_{\max}$ and $w_{\min}$, respectively, denote the maximum and minimum values of inertia weight.

3.2. MPSO

To avoid local optimization of PSO and accelerate the global searching ability of PSO, this paper introduced the nonlinear dynamic inertia weight coefficient in the inertia weight modified by PSO and proposed a modified particle swarm algorithm (MPSO).

4. MPSO-FSVM-based water quality prediction and evaluation of Chaohu Lake

4.1. Objective function

On the premise that the optimal water quality evaluation result is guaranteed, use MPSO to optimize the Γ and b, parameters of FSVM, so as to achieve the self-adaptive selection of FSVM parameters. If the actual water quality type at tth moment is $y(t)$ and the predicted water quality type is $\hat{y}(t)$, the difference between the actual water quality type $y(t) - \hat{y}(t)$.

As for the evaluation on water quality types, the actual water quality type data is n; use MPSO to optimize Γ and b, parameters of FSVM so as to minimize the quadratic sum of difference between the actual water quality type with FSVM and predicted water quality type.

4.2. Steps of algorithm

The steps of water quality evaluation with the FSVM optimized by MPSO are as follows:

Step 1: Normalize the water quality sample data. Establish training samples and test samples.

Step 2: Set the population size (*popsize*), learning factors c_1 and c_2 and maximum iterations $\max_{\text{gen}}$ of MPSO.

Step 3: Input the established training samples into the FSVM; calculate the functional value of particle fitness according to the objective function formula (7) to search for the individual and global optimal particle position and optimal value.

Step 4: Update the particle speed and position.

Step 5: Calculate the fitness and update position and speed simultaneously.

Step 6: If $\max_{\text{gen}}$, save the optimal solution. Conversely, if $\text{gen} = \text{gen} + 1$ and go to Step 4.

5. Experimental analysis

5.1. Water quality evaluation indexes

Water quality evaluation is to calculate and determine the water quality grade of sampling water samples through a certain mathematical model according to the water quality evaluation standard and all indexes of sampling water samples. There are numerous indexes for water quality analysis. Based on the quality standards for surface water environment, this paper used 6 water quality evaluation indexes including ammonia nitrogen, dissolved oxygen (DO), chemical oxygen demand (COD), permanganate index, total phosphorus and total nitrogen; their corresponding water quality grades are listed in Table 1.

Table 1. Water quality grades and content standards

Type	Type 1	Type 2	Type 3	Type 4	Type 5
Ammonia nitrogen (mg/l) $\leq$	0.15	0.50	1.0	1.5	2.0
DO (mg/l) $\geq$	7.5	6.0	5.0	3.0	2.0
COD /(mg/l) $\leq$	15	15	20	30	40
Permanganate index (mg/l) $\leq$	2.0	4.0	6.0	10	15
hline Total phosphorus (mg/l) $\leq$	0.02	0.10	0.20	0.30	0.40
Total nitrogen (mg/l) $\leq$	0.20	0.50	1.0	1.5	2.0

5.2. Data sources

Water samples of Chaohu Lake were collected as objects of evaluation of water quality. The sampling water intakes were respectively the intersection between Nanfei River and Chaohu Lake, the intersection between Paihe River and Chaohu Lake and dam outlet of Chaohu River. The longitudes and latitudes of sampling sites are listed in Table 2. The sampling time of water quality of Chaohu River was from 2010 to 2015. The sampling frequency was once per quarter. The change tendencies of all indexes of sampling water are shown in Figs. 1 and 2.

Table 2. Longitudes and latitudes of sampling sites

S/N	Sampling site	Longitude	Latitude
1#	Intersection between Nanfei River and Chaohu Lake	$117°24'40''$	$31°42'15''$
2#	Intersection between Paihe River and Chaohu Lake	$117°18'15''$	$31°41'30''$
3#	Dam outlet of Chaohu River	$117°51'46''$	$31°34'18''$

5.3. Diagram for spatial distribution of water quality

To visually observe the relations among various water quality indexes, the diagrams for spatial relations among all water quality indexes were drawn and are shown in Fig. 3.

5.4. Empirical results

To verify the validity and reliability of algorithm proposed in this paper, the water quality data of three sampling sites in Chaohu Lake from 2010 to 2015 were used as objects of research.

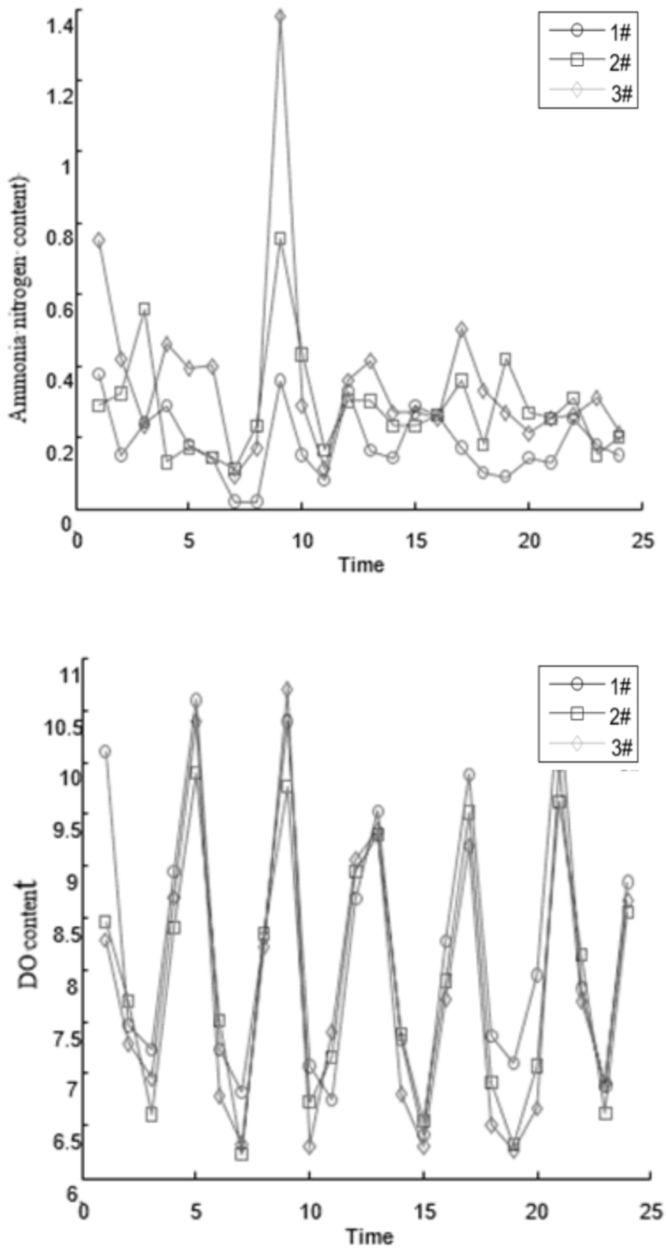

Fig. 1. Tendency chart for changes in water index data: up–Amonnia nitrogen content (mg/l), bottom–DO content (mg/l)

The parameter setting of MPSO is shown below. The population size is 20, the

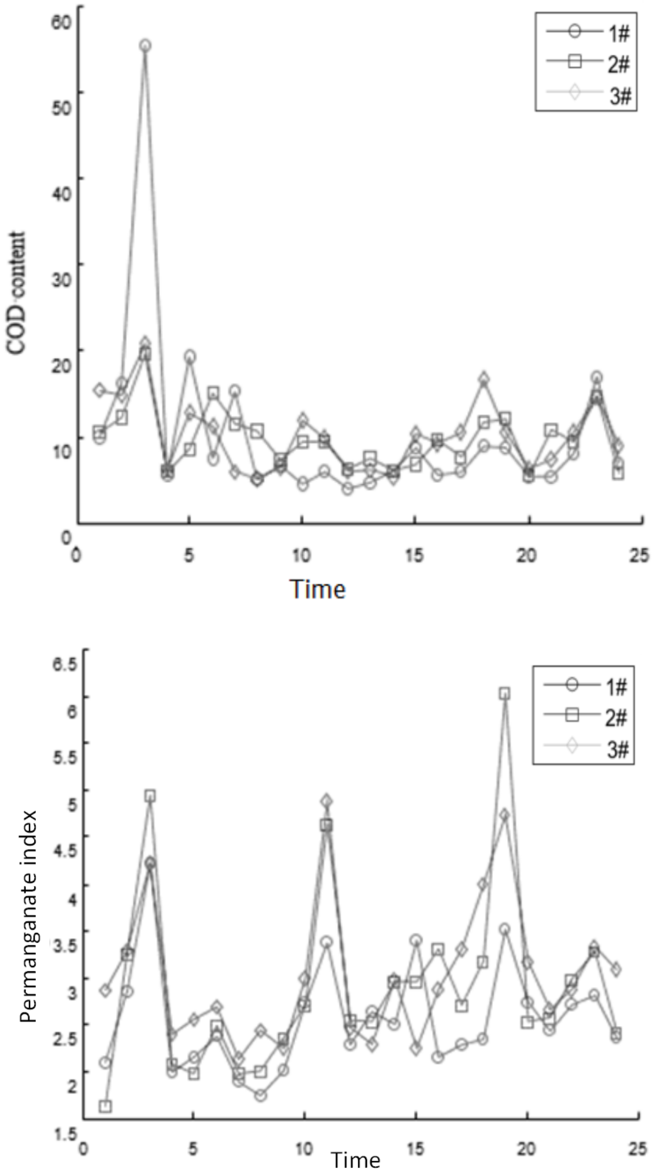

Fig. 2. Tendency chart for changes in water index data: up–COD content (mg/l), bottom–permanganate content (mg/l)

maximum iterations is 100, $c_1 = c_2 = 0.5$, $V_{\max} = 5$, $V_{\min} = -5\,\%$. The prediction and evaluation results of water quality of Chaohu Lake with the FSVM optimized by MPSO are shown in Figs. 4–5 and Table 3 and Table 4.

It can be seen from Table 3 that MPSO-FSVM, PSO-FSVM and FSVMall have higher accuracies in prediction and evaluation of water qualities of Chaohu Lake than SVM.

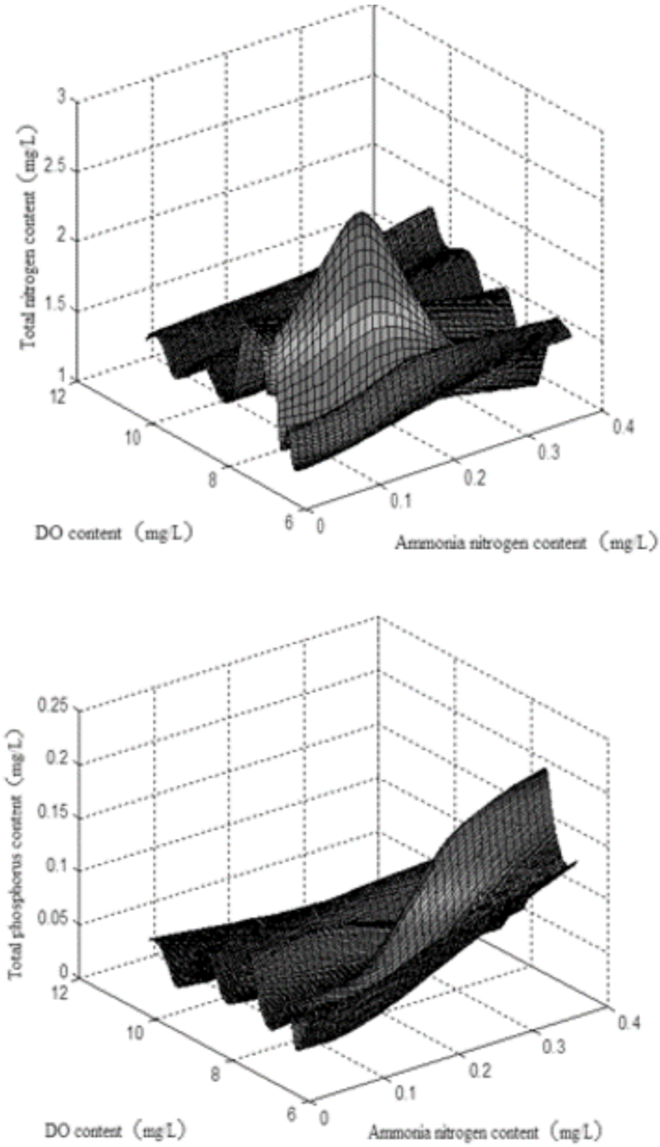

Fig. 3. Diagram for spatial relations among 1#, 2# and 3# sampling sites

There is no big difference in the accuracy rate of water quality grade classification between MPSO-FSVM and PSO-FSVM; however, PSO may easily lead to local optimization and has a poor stability; MPSO, which as strong global optimization ability, can effectively avoid the problem of local optimization. In addition, even though very high prediction accuracy can also be realized with FSVM, this method is time-consuming and inefficient; besides, it is required to set the searching scope in advance with this method, so it is difficult to control it and to guarantee the accuracy.

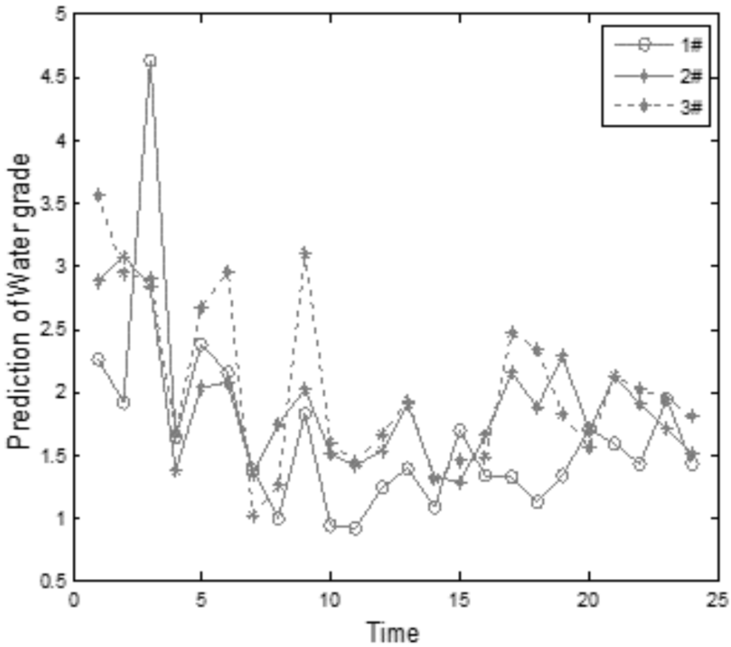

Fig. 4. Water evaluation results

Table 3. Evaluation results of water quality of Chaohu Lake

Time	2010.1	2010.2	2010.3	2010.4	2011.1	2011.2	2011.3	2011.4
1#	3	4	5	3	4	3	2	2
2#	4	4	3	3	3	3	2	3
3#	4	4	3	3	4	3	2	2

Table 4. Comparison of evaluation results of different water qualities

Method	MPSO-FSVM	PSO-FSVM	FSVM	SVM
Quantity of erroneous judgments	3	4	6	9
Accuracy	95.38 %	93.85 %	90.76 %	86.15 %

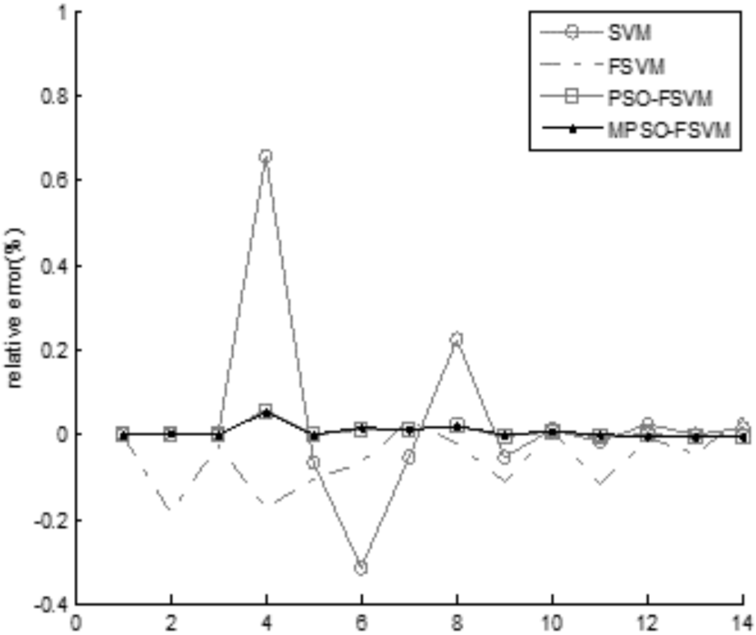

Fig. 5. Error comparison of different evaluation methods

6. Conclusion

In consideration of the low prediction accuracy and poor applicability in traditional water prediction and evaluation models, to realize high-accuracy water quality prediction and evaluation, a water quality prediction and evaluation model based on FSVM optimized by MPSO was put forward on the basis of FSVM. The MPSO was used to optimize Gamma and b, parameters of FSVM and realize the high-accuracy water quality prediction and evaluation under the condition of optimal parameters. The monitoring data of water quality of Chaohu Lake from 2010 to 2015 were selected as objects of research; the water quality prediction and evaluation of Chaohu Lake were achieved based on studying the water quality change trend and time-space relationships. It is found through comparison of 4 types of water quality evaluation and prediction methods including MPSO-FSVM , PSO-FSVM , FSVM and SVM that the prediction accuracy of MPSO-FSVM reaches up to 95.38 %, which is higher than the prediction accuracies of PSO-FSVM , FSVM and SVM. Therefore, it indicates that MPSO-FSVM has a higher precision and adaptability in water quality evaluation and prediction; and it also has a better effect.

References

[1] H. B. Xue, Y. Wei: *Optimize an optimal GM(1,1) based on the discrete function*

with exponential law once again. Mathematics in Practice and Theory *1* (2009), No. 1, 242–246.

[2] X. H. Yao, M. R. Fei, K. Li, H. Kong, B. Zhao: *Recognition of blue-green algae in lakes using distributive genetic algorithm-based neural networks.* Neurocomputing *70* (2007), Nos. 4–6, 641–647.

[3] X. Liu, F. Dong, G. He, J. Liu: *Use of PCA-RBF model for prediction of chlorophyll-a in Yuqiao Reservoir in the Haihe River Basin, China.* Water Science & Technology: Water Supply *14* (2014), No. 1, 73–80.

[4] U. Natarajan, V. M. Periasamy, R. Saravanan: *Application of particle swarm optimisation in artificial neural network for the prediction of tool life.* The International Journal of Advanced Manufacturing Technology *31* (2007), Nos. 9–10, 871–876.

[5] J. G. Yang, S. Y. Wen, H. Zhou, K.z,F. Cen: *An optimized BP network model using genetic algorithm for predicting the ignition-stability index of pulverized coal.* Journal of Power Engineering *26* (2006), No. 1, 81–83.

[6] K. P. Wu, S. D. Wang: *Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space.* Pattern Recognition *42* (2009), No. 5, 710–717.

[7] H. Yoon, S. C. Jun, Y. Hyun, G. O. Bae, K. K. Lee: *A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer.* Journal of Hydrology *396* (2011), Nos. 1–2, 128–138.

[8] D. KarabogaD, B. Akay: *A survey: Algorithms simulating bee swarm intelligence.* Artificial Intelligence Review *31* 2009), Nos. 1–4, 61–85.

[9] H. S. Chang: *Converging marriage in honey-bees optimization and application to stochastic dynamic programming.* Journal of Global Optimization *35* (2006), No. 3, 423–441.

[10] B. Parinet, A. Lhote, B. Legube: *Principal component analysis: An appropriate tool for water quality evaluation and management—application to a tropical lake system.* Ecological Modelling *178*, (2004), Nos. 3–4, 295–311.

[11] L. C. Hsu: *Applying the Grey prediction model to the global integrated circuit industry.* Technological Forecasting and Social Change *70* (2003), No. 6, 563–574.

[12] F. Moatar, M. Meybeck: *Compared performances of different algorithms for estimating annual nutrient loads discharged by the eutrophic river Loire.* Hydrological Processes *19* (2005), No. 2, 429–444.

[13] W. Yan, S. Zhang, P. Sun, S. P. Seitzinger: *How do nitrogen inputs to the Changjiang basin impact the Changjiang River nitrate: A temporal analysis for 1968–1997.* Global Biogeochemical Cycles *17* (2003), No. 4, 1091–1100.

[14] M. Aminian, F. Aminian: *Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor.* IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing *47* (2000), No. 2, 151–156.

[15] F Aminian, M. Aminian: *Fault diagnosis of analog circuits using bayesian neural networks with wavelet transform as preprocessor.* Journal of Electronic Testing *17* (2001), No. 1, 29–36.

Received July 12, 2017