

Article

Water Quality Prediction Model of a Water Diversion Project Based on the Improved Artificial Bee Colony–Backpropagation Neural Network

Siyu Chen *, Guohua Fang *, Xianfeng Huang and Yuhong Zhang

College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China; hxfhuang2005@163.com (X.H.); 17314978694@163.com (Y.Z.)

* Correspondence: 17802673959@163.com (S.C.); hhufgh@163.com (G.F.)

Received: 1 May 2018; Accepted: 16 June 2018; Published: 18 June 2018



Abstract: Prediction of water quality which can ensure the water supply and prevent water pollution is essential for a successful water transfer project. In recent years, with the development of artificial intelligence, the backpropagation (BP) neural network has been increasingly applied for the prediction and forecasting field. However, the BP neural network frame cannot satisfy the demand of higher accuracy. In this study, we extracted monitoring data from the water transfer channel of both the water resource and the intake area as training samples and selected some distinct indices as input factors to establish a BP neural network whose connection weight values between network layers and the threshold of each layer had already been optimized by an improved artificial bee colony (IABC) algorithm. Compared with the traditional BP and ABC-BP neural network model, it was shown that the IABC-BP neural network has a greater ability for forecasting and could achieve much better accuracy, nearly 25% more precise than the BP neural network. The new model is particularly practical for the water quality prediction of a water diversion project and could be readily applied in this field.

Keywords: BP neural network; Artificial Bee Colony Algorithm; water quality; prediction; water diversion project

1. Introduction

Water shortage and water resource pollution have become major problems in China. In order to solve the problem of water resource imbalance, water diversion projects in many areas have been constructed. Water quality is the key to the success of a water diversion project. The prediction of water quality is to predict the variation trend of water environment quality at a certain time in the future. The water quality prediction is of great significance to the planning and control of water quality. In order to make the plan for water pollution prevention and control, it is necessary to predict the changes of water quality at different pollution levels in the future so as to formulate a reasonable plan. For a water diversion project, it is more important to predict the water quality because quite a significant amount of the water is transferred for solving daily drinking problems. Therefore, it is of great significance to explore the methods of water quality prediction in the present society.

For the development of a water quality model, the following three stages can be summarized [1]: (1) From the mid-1920s to the 1970s, the models mainly focused on the study of oxygen balance and belonged to the one-dimensional steady-state model. (2) From the 1970s to the mid-1980s, the water quality model developed rapidly. The state variable (water quality component) was increased, the multi-media environment ecological synthesis model was established; the water dynamics model was incorporated into the multi-dimensional model system; the one-dimensional steady-state model

was developed into a multi-dimensional dynamic model. (3) From the mid-1980s to the present, the attention of scientific researchers gradually shifted to improving the reliability and evaluation capacity of water quality models. Not only concerning the mechanism water quality model but also many new techniques and methods appeared.

Although the mechanism model takes into account the physical, chemical, and biological factors that result in the changes of water quality, these models are relatively complex, the required water quality data is very large and these factors limit the further application of the model to the water to some degree [2]. In recent decades, with the development of computer technology, the non-mechanism model has become a hotspot for research on the water quality prediction model, such as the Markov method [3], the grey prediction model method [4], the multiple regression model, the Artificial Neural Network (ANN) prediction model method, the Support Vector Machine Model [5], the Genetic Algorithm (GA), and wavelet analysis. The characteristic of these methods is to establish a water quality prediction model with a certain algorithm from the perspective of the variation in water quality data and without considering the relationship of the water pollution and the changing mechanism. In other words, the modeling method is a kind of black box type.

The water environment system is a system with strong nonlinear and nondeterministic characteristics. The traditional linear prediction model cannot fully reflect its changing regulation and the prediction accuracy also cannot now satisfy requirements. The water pollution process is so complex that it is not only affected by natural factors and pollutant discharge, but also by factors such as social and economic development, resulting in the nonlinear water environment system [6]. Now many of the reaction mechanisms of a chemical, biological process cannot be expressed exactly by a mathematical equation which limits the applicability and accuracy of the traditional water quality mathematical model.

ANN with the characteristics of the cranial nerve system is a suitable way to solve nonlinear and uncertain problems, and the application of the ANN model in water quality prediction has developed more widely and strongly. The BP neural network is the most common and universal form of ANN which has a strong nonlinear mapping capability and a flexible network structure. The extreme learning machine (ELM), as a kind of machine learning algorithm designed for feedforward neural networks, was put forward in 2004 by Huang, Zhu, and Chee. The appearance together with its motivation is to overcome the low learning efficiency and complicated parameter setting issues of traditional neural networks algorithm. Although fixed extreme learning machines, such as ELM, OS-ELM, are recognized as neural network training methods with the fastest learning speeds, they are still controversial in their robustness and stability. Thus, the BP neural network was chosen to be the basic method of this article to ensure robustness and stability [7]. However, because the BP neural network has some defects like slow convergence speed, easily falling into a local minimum value while the network is sensitive to the initial value, many scholars have put forward various improved BP algorithms, for example, GA and Particle Swarm Optimization (PSO) Algorithms have been used to improve the BP neural network water quality prediction model. Maier [8] introduced how to use the BP neural network to predict future water quality parameters on the basis of a large number of historical data. Cao [9] established the BP neural network water quality prediction model in the Danjiangkou Reservoir and predicted the water quality indices in the Danjiangkou Reservoir area; while, the relative error between the actual detection value and the predicted value was less than 7%. Chang [10] established a dynamic, timely, and advanced BP neural network prediction model, with short time and high accuracy in water quality prediction. Chen [11] used a neural network of variable structure to predict the water quality parameters and obtained better results. Singh [12] proposed an ANN model whose computed values of DO and BOD were in close agreement with their respective measured values in the river water. Zheng [13] proposed that the immune PSO algorithm which was used to train the network of hidden layer nodes and the weights of the network, as well as to forecast the sewage effluent water quality, has an ideal forecasting effect. Gao [14] optimized the BP neural network by improving the input variables through grey correlation analysis. Zhang [15]

combined GA with the neural network to construct the genetic neural network prediction model of water quality time series to improve the stability of the prediction results. Wang [16] for the first time verified that the GA-GRNN model is efficient for water quality prediction under normal conditions and can be used to ensure the security of water delivery and water quality in the South-to-North Water Diversion (SNWD) Project. To sum up, the population-based meta-heuristic algorithms such as the GA, PSO and ABC algorithm can efficiently speed up the search process [17]. However, the ABC algorithm which has been proved by Karaboga, the proposer of the ABC algorithm, to be more precise than PSO and GA has not been applied widely in water quality prediction [18–20], let alone in the prediction of water diverse projects.

In this study, the Eastern Route (ER) Project of South-to-North Water Diversion Project was selected as an example, focusing on cascade pumping stations, water diversion projects, and water environment character such as complicated dynamic conditions, and an IABC-BP neural network model was constructed. Equipped with the generalization and mapping capability of the neural network and the global iteration and local search capability of the ABC algorithm, this IABC-BP neural network modeling method was proposed for using the monitoring indices of the first stage to forecast the indicators of the sixth stage. Based on the multivariate correlation between the water quality of the first and sixth stage and the equilibrium principle of mixtures, the model represents the process of how the variation of water quality of the first stage influences the water quality of the sixth stage, and the essence of this model is the study and calculation of water quality movement evolution law in water.

This study is organized as follows: Section 2 provides the materials and methods used for this study, i.e., the study area, study data, and the principles of the BP network and ABC algorithm. Section 3 describes the research design and process, i.e., the improvement of the ABC algorithm and the experiment setting such as the initialization of the parameters of the IABC-BP neural network. Section 4 provides the application results, with corresponding analyses and discussions. Finally, the conclusions and further research direction are presented in Section 5.

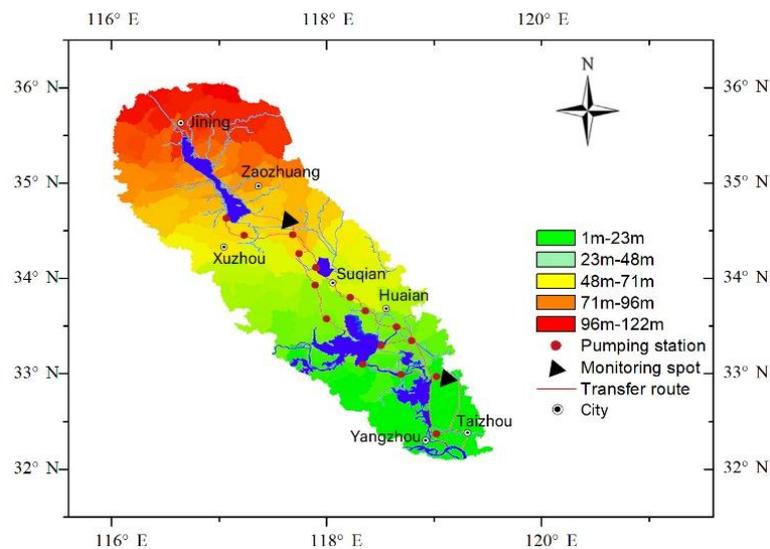
2. Research Methods

2.1. Research Area

The ER Project of the SNWD Project in China starts from Yangzhou City of Jiangsu Province and transfers the water of the downstream Yangtze River to Tianjin by raising the water in stages through the Beijing-Hangzhou Grand Canal and the watercourse parallel to it. It alleviates water shortages along the way, including North Jiangsu, Anhui, Hebei, Tianjin, and even for Beijing in the near future. The total extent of the ER is 1466.5 km. The project was officially brought into operation in 2013 while there was severe water contamination before. Thus, water pollution, the problem which has always been ignored, was finally put on the agenda. Therefore, at the end of 2013, the supplied water along the east route reached the standard of drinkable water, the third type of surface water. However, the water along the riverbank is still plagued by living sewage and industrial contamination, leading to dead fish and unpleasant conditions for living and tourism to some degree, and this impact can be reflected by the prediction model according to Table 1. In the Table 1 DO represents dissolved oxygen, COD_{Mn} represents permanganate index, and BOD₅ represents the biochemical oxygen demand of five days. The classification can be decided by the model results, and if the classification of the sixth stage is worse than the first stage, it illustrates that there could be pollution that affects the water quality along the transfer route. As shown in Figure 1, the study area is from the first stage to the sixth stage pumping station in Jiangsu Province, and the sixth pumping station at the border of Jiangsu and Shandong is the starting point of the intake area.

Table 1. Surface water quality standard of PRC (GB 3838-2002).

Classification	pH (Non-Dimensional)	DO \geq (mg/L)	CODMn \leq (mg/L)	NH ₃ -N \leq (mg/L)	Petroleum \leq (mg/L)	Volatile Phenol \leq (mg/L)	BOD ₅ \leq (mg/L)
I	6~9	7.5	2	0.15	0.05	0.002	3
II	6~9	6	4	0.5	0.05	0.002	3
III	6~9	5	6	1.0	0.05	0.005	4
IV	6~9	3	10	1.5	0.5	0.01	6
V	6~9	2	15	2.0	1.0	0.1	10

**Figure 1.** Experimental area in Jiangsu Province.

2.2. Research Data

In order to prevent water pollution and ensure the supplied water quality, two measurement locations were selected for the analysis. One location is at the first stage of the pumping station, in Baoying County of Yangzhou City near the origin of the east route, downstream of the Yangtze River. The other which is located in Shantou is at the sixth stage, around the starting point of the intake area at the boundary of Jiangsu and Shandong Province. The datasets of both the first and sixth stages were measured by Jiangsu Province Hydrology and Water Resources Investigation Bureau, comprising seven water quality parameters monitored when the pumping stations were operating in 2016. Unlike the river, the water diversion project does not operate every day, it only operates when the intake areas are dry, and the operating period usually lasts for at least five days at a time. Besides, it takes a whole day to raise water from the first stage to the sixth stage. Thus, in this study, we used the first day's data of the first stage to predict the data of the sixth stage of the next day. In the present study, three main parameters were selected as neurons of the input layer, they are DO, COD_{Mn}, and tBOD₅. The data of these parameters which are available for analysis are on a daily basis for a period of about 2 months, there are four periods of water transfer, 1–28 January, 2–15 March, 27 March–9 April and 21 April–10 May. Thus, the corresponding periods of the monitored data of the sixth stage are one day later than the periods mentioned above. According to the principles of accuracy, representativeness, and statistics, the training samples in this study should cover most of the possibilities which means they should occur in all the seasons that the pumping stations operate. They represent most of the cases that may occur and also the number of the training samples should be as many as possible. Thus, in this study, the water quality data of the first 50 days for DO, COD_{Mn}, and BOD₅ were used for model training, because almost all of the possibilities were covered by the first 50 sets. The remaining 16 sets of data were used for verification of the model prediction results. The statistical properties of the water quality time series data are demonstrated in Table 1. The maximum, minimum, mean value,

standard deviation, skewness, and kurtosis describe the variability of those parameters. As depicted in Table 2, DL represents detection limited, the potential of hydrogen (pH), ammonia nitrogen (NH₃-N), content of petroleum, and volatile phenol have low skewness coefficients. On the other hand, DO, COD_{Mn}, BOD₅ have comparatively higher skewness coefficients which indicates that there may be a difference between the median and mean value of these variables. It is probable that there are a few extreme values existing to affect the mean value of the parameters.

Table 2. Statistical properties of the water quality parameters.

Parameters	Max	Min	Mean	Std. Dev.	Skewness	Kurtosis
Water temperature (°C)	19.00	6.80	14.29	4.3245	0.0048	−1.8145
Potential of Hydrogen	8.21	7.76	7.99	0.1135	0.0267	−0.9186
Dissolved oxygen (mg/L)	12.40	6.30	9.19	1.5658	−0.2266	−0.8880
Permanganate index (mg/L)	6.00	3.00	4.81	0.6107	−0.2669	0.6001
Biochemical Oxygen Demand of five days (mg/L)	3.70	0.23	2.91	0.4986	−0.1729	−1.2901
Ammonia nitrogen (mg/L)	0.93	1.90	0.49	0.1206	0.0223	2.6669
Content of petroleum (mg/L)	<DL	<DL	-	-	-	-
Content of volatile phenol (mg/L)	<DL	<DL	-	-	-	-

2.3. Principle of Backpropagation Neural Network

The BP neural network is the most popular model among all the neural network architectures that are based on the research of biological neurons. The basic structure of the BP neural network is a feedforward network which is composed of one input layer, several hidden layers, and one output layer. Neurons between every two layers connect each other by connection weights and there is no relation between neurons in the same layer. The output of the former layer is the input of the next layer.

The BP neural network consists of forward propagation of the signal and back propagation of the error. In the first stage, the sample indices enter the input layer and then are transmitted to the output layer after being processed layer by layer. If the actual output values are inconsistent with the desired values, then the algorithm turns into the phase of back propagation. In this stage, the output values are transmitted layer by layer back to the input layer in a certain way and the error is allocated to each neuron of the network so that each error signal which is a prerequisite of revising the weight value between every two neurons is obtained. Its typical structure is shown in Figure 2.

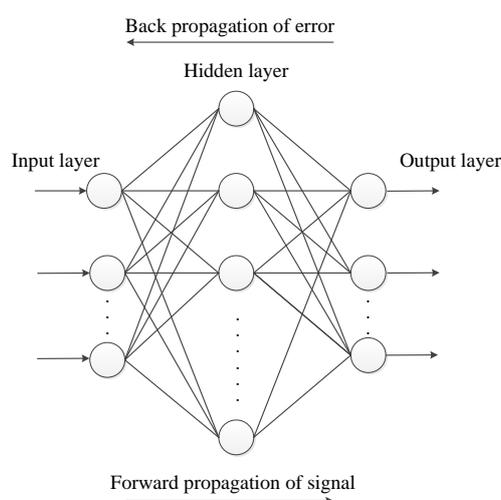


Figure 2. Architecture of the typical BP neural network.

The BP algorithm has the ideal training effect in the neural network within five layers, but is difficult to train in the deep structure because of the local minimum problem in the non-convex objective function. In addition, the BP algorithm also has the problem of easily converging to local minimum values, resulting from the training, and often caused by the initialization of random values from the optimal region. Therefore, in order to obtain a more accurate water quality prediction model, the BP algorithm needs to be improved. For the former, the Kolmogorov theorem shows that as long as the number of neurons of hidden layer nodes is appropriate, even with only one hidden layer of the three-layers, the BP neural network can realize arbitrary nonlinear function approximation, so this article only sets one hidden layer in order to reduce unnecessary calculation; while for the latter, because the ABC algorithm has strong robustness and fast convergence speed it also can effectively avoid the BP neural network getting into a local optimum. Besides, compared with the GA and PSO algorithm, the ABC algorithm has better optimization performance, thus the ABC algorithm was selected to optimize the BP neural network in this study [21].

2.4. Principle of the Artificial Bee Colony Algorithm

The ABC algorithm is a novel combinational optimization algorithm with meta-heuristic intelligence which was proposed by the team of Dervis Karaboga in 2005. The ABC algorithm which was inspired by the intelligent behavior of a honey bee swarm when foraging for food has been gaining increasing popularity over the last decade. The ABC algorithm has become a highly efficient method to solve complex nonlinear optimization problems and it has been proved that it outperforms GA, PSO, Particle Swarm Inspired Evolutionary Algorithm (PS-EA), and some other population-based algorithms in optimizing multivariable functions. It also can be efficiently used for solving engineering problems with high dimensionality. In this study, there are over 100 variables to be optimized, thus ABC is more suitable for this work.

There are three types of bee groups in the ABC algorithm: employed bees, onlookers, and scouts and each individual bee has its own work and should cooperate with other bees. There is only one employed bee for one food resource, in other words, food number equals the quantity of the employed bees. Each position of nectar source represents a possible solution of the optimization problem and the quality of this nectar shows the fitness of this solution. There are three phases of the ABC algorithm that correspond to the three types of bees: phase of employed bees, onlookers, and scouts. In the first stage, the position of the nectar source is abstracted into a point of the solutions and the employed bees start to forage in the neighborhood during the iteration process. In the next phase, roulette is used to employ the onlookers for foraging anew a nectar according to the quality of the previous nectar, namely fitness. At last, the nectar resource is abandoned if it cannot be improved by updating time after time and the employed bees become scouts again for foraging new food resource randomly. The above, employed bees are for maintaining a solution with good quality, onlookers are for improving the speed of convergence and the scouts are for strengthening the ability to get rid of the local optimum.

2.5. ABC-BP Neural Network Model

The ABC algorithm is used to search for the optimum connection weight values between neurons and the threshold values of each neuron in this new model which is equipped with the abilities of local searching and generalization from both the ABC algorithm and BP neural network.

The specific steps of the ABC-BP Neural Network are listed as the following:

Step 1: Establish the BP neural network:

Determine the numbers of hidden layers, neurons in the hidden layer j , ($j = 1, 2, \dots, N_h$), input layer nodes i , ($i = 1, 2, \dots, N_i$), output layer nodes k , ($k = 1, 2, \dots, N_o$), and training samples p , ($p = 1, 2, \dots, N_s$). Then create the objective function whose result needs to be optimized according to the conditions above.

Step 2: Initialize the parameters of the ABC algorithm:

Initialize the size of the bee colony N_c , the numbers of the employed bees N_e , onlookers N_o , and food sources N_f , the times of iteration t ($t = 1, 2, \dots, MCN$), the maximum of the trials for foraging new nectar tri ($tri = 1, 2 \dots$ limit) and other end conditions of the algorithm. The food sources X^t are generated according to Formula (1).

$$X_i^t = L + rand(-1, 1) \times (U - L) \quad (1)$$

In Formula (1), L and U are the lower and upper limit values of X_i^t ($i = 1, 2, \dots, N_f$) in the searching area. $X^t = x_{ij}^t$ ($j = 1, 2, \dots, D$), D is the dimension of each solution and the D dimensioned vector X_i^t represents the connection weight and threshold values of the network in Formula (1), while the dimension D satisfies the Formula (2).

$$D = (N_i + 1) \times N_h + (N_h + 1) \times N_o \quad (2)$$

Step 3: Calculate the fitness value of each nectar source according to Formula (3).

$$fit_i^t = 1 / (1 + f(X_i^t)) \quad (3)$$

In Formula (3), $f(X_i^t)$ is the value of the majorized objective function of X_i^t and the value in the ABC-BP neural network is always the mean square error, thus $f(X_i^t)$ is always more than or equal to 0. Therefore, the ideal state is when the fitness reaches 1.

Step 4: Update the new nectar source by the employed bees:

The employed bees forage for the new food source in the neighborhood of nectar source X_i^t at the beginning of this stage according to Formula (4). When all of the employed bees finish the computation of Formula (2), they fly back to the information sharing area.

$$v_{ij}^t = x_{ij}^t + rand(-1, 1) \times (x_{ij}^t - x_{dj}^t) \quad (4)$$

In Formula (4), $i \neq d$, it presents that a nectar source which is not the same as solution i is picked. When the fitness of the new nectar source V_i^t is better than the fitness of X_i^t , the new solution replaces the previous one by the greedy method. Otherwise, the food source remains as the previous value and the number of failures plus one $tri = tri + 1$.

Step 5: Update the new nectar source by onlookers:

Onlookers decide which employed bee they will follow according to Formula (5) by roulette strategy. Then, onlookers turn into employed bees and find a new food source around X_i^t according to Formula (4) and choose a better nectar source by the greedy method.

$$p_i^t = fit_i^t / \sum_{i=1}^n fit_i^t \quad (5)$$

In Formula (5), p_i^t is the probability that the onlookers choose the nectar source X_i^t .

Step 6: Determine if there is a solution abandoned:

During the searching period, if the bee colony cannot find a better nectar source around nectar source X_i^t until the limit trial time is approached, X_i^t is abandoned. The corresponding employed bee becomes scout for researching another new food source X_i^{t+1} to replace X_i^t according to the upper formula of Formula (6) and the new round starts from step 4. Otherwise, record the best solution

hitherto and determine whether the end conditions can be met, if so, print the optimal solution and the process ends; if not, start again from step 4. The process above was abstracted into Formula (6).

$$X_i^{t+1} = \begin{cases} L + rand(0,1) \times (U - L) & tri \geq limit \\ X_i^t & tri < limit \end{cases} \quad (6)$$

Step 7: Transform the optimal solution into the connection weight values and threshold values of the BP neural network, then simulate and test the neural network by the solution.

3. Research Design and Process

3.1. An Improved Artificial Bee Colony Algorithm

In this study, an IABC-BP model that combines an improved ABC algorithm with a BP neural network was proposed to achieve the goal, that is, predicting the water quality more accurately than the traditional BP neural network model and ABC-BP neural network.

3.1.1. Defects of the ABC Algorithm

In the process of optimization, the ABC algorithm has a good optimization effect. Due to less control parameters, good-quality parallelism, strong robustness, simple operation and also because it is easy to implement, it has attracted the attention of many experts and scholars at home and abroad. Although the ABC algorithm has many advantages in dealing with optimization problems, the algorithm still has some defects such as slow convergence speed and local optimal solution.

3.1.2. Improvement of ABC Algorithm

This study analyzed and improved the above two defects of the ABC algorithm.

Analysis of slow convergence speed and the corresponding improvement:

The initial solution is the starting point for the algorithm to search, and the distribution of the individuals in the initial group is significant for the searching parameters of the algorithm. The basic ABC algorithm adopts the random generation method which randomly generates several individuals to form the initial group. The shortcoming of the method is that it is difficult to control the distribution of the initial solution, what is more, it cannot guarantee the group to be distributed uniformly within the solution space. It is more likely that the spatial distribution of the initial population would be in the same dense area, this results in the fact that the population diversity might be decreased and finally may reduce the convergence rate of the algorithm. Therefore, if the initial population generated during the initialization process is not reasonable, it would have a great impact on the performance of the algorithm.

In order to reduce the influence of the initial solution and make sure that the initial solution can be distributed uniformly in the solution space, it is necessary to improve the generation strategy of the initial group. Zhang [22] first proposed the generation of initial solutions using the inter-cell generation method and applied it for the improved application of GA; Bao [23] applied the inter-cell generation method to the ABC algorithm, the experimental results showed that the improvement is effective. In this study, the random generation method of the basic ABC algorithm was replaced by the inter-cell generation method to generate the initial solution.

The inter-cell generation method refers to dividing the scope of each parameter to be optimized of the optimization problem into N_f ranges, and then one individual is generated in each area so that the whole initial group is formed. Therefore, the new initial solution generation process was abstracted into Formula (7) to replace Formula (1).

$$x_{ij}^t = (i \times lb + rand(0,1) \times (ub - lb)) / N_f \quad (7)$$

In Formula (7), the meanings of i, j , and N_f are the same as Formula (1). ub and lb are the upper and lower limit values of x_{ij}^t . This formula ensures that the initialized solutions are distributed uniformly.

Analysis of local optimum and the corresponding improvement:

The basic ABC algorithm has two renewal processes: the updating process of the employed bees and onlookers. The former is a process of global searching, focusing on developing and probing the searching space to find the possible optimal area. Its main purpose is to maintain the diversity of the group. The latter is a local searching process, focusing on making full use of the effective information in the community to exploit searching space, which guides the algorithm to place an emphasis on those individuals with higher fitness values. In practice, development and exploitation are contradictory. In order to improve the optimization ability of the algorithm, there must be an effective trade-off between them. In the basic ABC algorithm, the two renewable processes are updated with Formula (4), and the new solution is generated by moving the old solution to or away from the randomly selected neighbor solution in the population. Parameters are generated randomly with Formula (4), randomness of the candidate solutions is large and this leads to the strong global searching ability of the employed bees but weaker local searching ability of onlookers so that at last it can reduce the optimal performance of the algorithm. Therefore, it is necessary to improve the updating formula in the ABC algorithm.

The PSO algorithm is also a popular swarm optimization algorithm based on population, and one of the advantages of the PSO algorithm is the fast-speed convergence. It adds the present optimal solution and the global optimal solution to the updating formulas in order to accelerate the convergence rate of the algorithm. Inspired by the PSO algorithm, the IABC algorithm in this study also introduced the present optimal solution and global optimal solution to guide the search of the solution.

The updated Formula (4) of the employed bees in IABC was changed to Formula (8):

$$v_{ij}^t = x_{gbest,j}^t + rand(-1, 1) \times (x_{r_1,j}^t - x_{r_2,j}^t) \quad (8)$$

The updated Formula (4) of the onlookers in IABC was changed to Formula (9):

$$v_{ij}^t = x_{pbest,j}^t + rand(-1, 1) \times (x_{r_1,j}^t - x_{r_2,j}^t) \quad (9)$$

In the formulas above, $x_{pbest,j}^t$ is the present optimal solution and $x_{gbest,j}^t$ is the global optimal solution. $i, r_1, r_2 \in \{1, 2, \dots, N_s\}$ and $r_1 \neq r_2 \neq i$. On one hand, $rand(-1, 1) \times (x_{r_1,j}^t - x_{r_2,j}^t)$ in Formulae (8) and (9) ensures the large randomness and strengthens the ability of global searching. On the other hand, $x_{gbest,j}^t$ and $x_{pbest,j}^t$ are added to improve the capability of local searching. The specific flow chart is described as Figure 3.

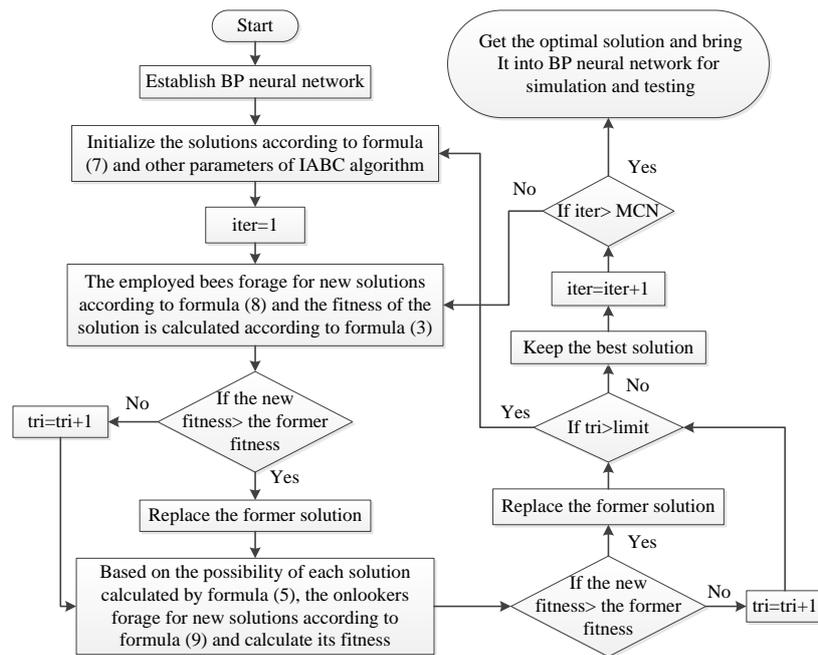


Figure 3. The flow chart of IABC-BP neural network.

3.2. Experiment Setting

The experiment in this study was divided into three parts. The first part was to compare the performance of the pure BP neural network and ABC-BP algorithm. The second part compared the performance of the BP and IABC-BP algorithm. The third part compared the performance of the IABC-BP and ABC-BP algorithm. The water quality of the water diversion project was predicted with the optimization of the connection weight and threshold value in the neural network.

3.2.1. Objective Function

Establish an unconstrained and nonlinear optimization problem of network mean square error for a certain network system. The optimization problem was abstracted into Formula (10):

$$\min \{f(X_i^t)\} X_i^t \in R^D \tag{10}$$

The transmission of the p^{th} sample: the output content of the input layer $o_{pi}^t = x_{pi}^t, (i = 1, 2, \dots, N_i)$, the input values of the hidden layer are obtained by the output values of the input layer $net_{pj}^t = \sum_{i=1}^{N_i} w_{ji}^t \times o_{pi}^t + \theta_j^t$; The output values of the hidden layer $o_{pj}^t = f(net_{pj}^t)$, the input values of the output layer are obtained by the output values of the hidden layer $net_{pk}^t = \sum_{j=1}^{N_h} w_{kj}^t \times o_{pj}^t + \theta_k^t$, the output values of the network $o_{pk}^t = f(net_{pk}^t)$. The actual observed values are represented by y_{pk}^t . Therefore, the objective function was established as Formula (11):

$$f(X_i^t) = \frac{1}{N_s} \sum_{p=1}^{N_s} \sum_{k=1}^{N_o} (y_{pk}^t - o_{pk}^t)^2 \tag{11}$$

3.2.2. Parameters Initialization of the IABC-BP Algorithm

The Kolmogorov theorem shows that as long as the number of neurons of hidden layer nodes is enough, even with only one hidden layer of the three-layers, the BP neural network can realize

arbitrary nonlinear function approximation, so this study set only one hidden layer in order to reduce unnecessary calculation. The research object was the water quality prediction data of the monitoring section of the sixth stage in the ER of the SNWD project. According to the three selecting principles of water quality evaluation parameters, principles of pertinence, proportionality, and feasible monitoring technology, and combining with the analysis of the monitoring data of the water quality, three main parameters were selected as neurons of the input layer, DO, COD_{Mn}, and BOD5. We extracted 66 monitoring datasets of these parameters according to 2.2. The data of the first 50 days for DO, COD_{Mn}, and BOD5 were used for model training and the remaining 16 sets of data were used for the verification of the model prediction results. Thus, we obtained the numbers of input and output node as 3 and after the times of the test as shown in Figure 4, the suitable number of neurons in the hidden layer is 7.

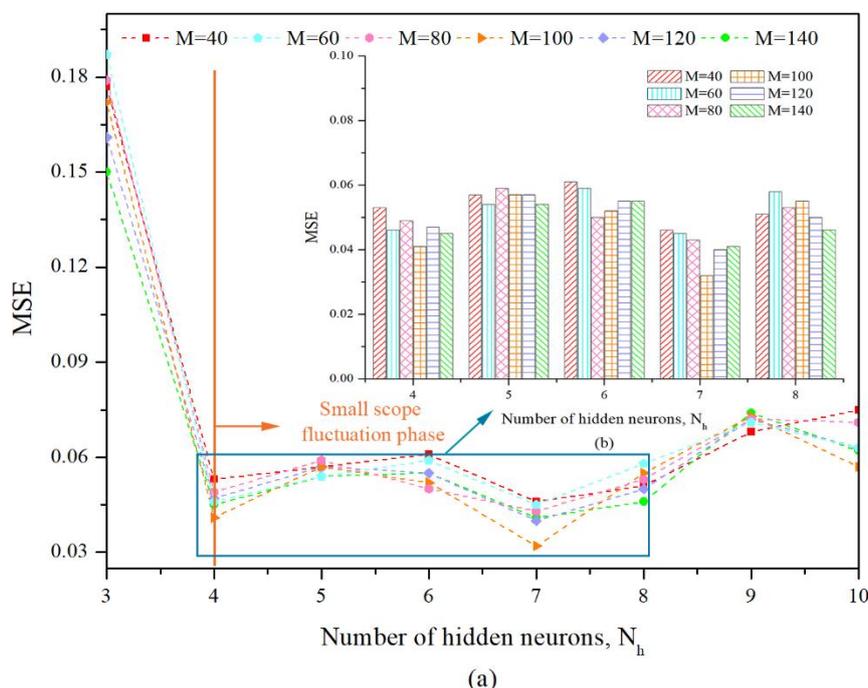


Figure 4. Performance of IABC-BP models with various parameter settings: (a) MSE distribution at the entire range (3–10) of N_h ; (b) MSE distribution at the selected range (4–8) of N_h .

According to Figure 4, it is obvious that the greater the swarm size s , the higher is the probability to acquire the optimum solution. However, the computation would be more complicated if the size of bee colony was too big. After repeated simulation tests, the size of the bee colony in this study was determined to be 100, the limit which should be more than the dimension D of each solution was decided to be 50, the MCN was 500 and the $N_e = N_o = N_f = 50$.

Therefore, a 3-7-3 IABC-BP model with a population size of 100 was adopted in this study to predict the water quality of the water diversion project. The 3-7-3 BP neural network was shown as Figure 5.

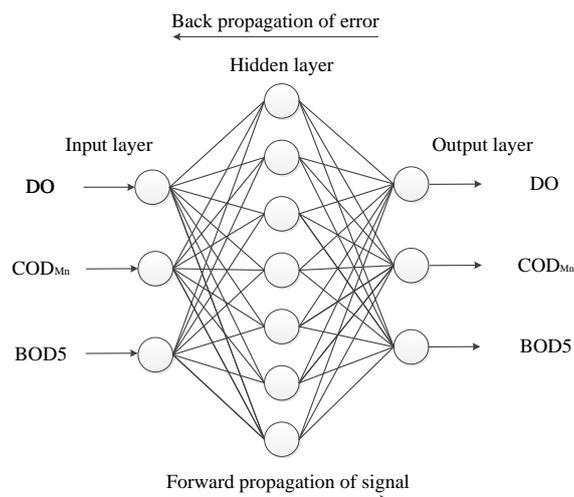


Figure 5. The 3-7-3 BP neural network model.

3.3. Result Verification

In order to ensure the accuracy of the models, three indicators were selected to testify the results.

3.3.1. Relative Error

Relative error is the ratio of an error in a measured or calculated quantity to the magnitude of that quantity.

3.3.2. Coefficient of Determination

The coefficient of determination is a measure used in statistical analysis that assesses how well a model explains and predicts future outcomes. It is indicative of the level of explained variability in the data set.

3.3.3. Nash–Sutcliffe Efficiency Coefficient

The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models.

4. Results and Analyses

4.1. Convergence Performance Analysis

The two defects of the IABC-BP model (slow convergence speed and local optimum) were improved, this impact is reflected by the performance of convergence. In this study, the fitness of the model reflects the quality of the result, thus, the variation of the fitness was chosen to depict the convergence performance of the model. The convergence performance of each model is demonstrated in Table 3 and Figure 6. Convergence accuracy and speed were analyzed to illustrate the convergence performance of each model.

Table 3. Performance of convergence of each model.

Iteration Times	Fitness _{BP}	Fitness _{PSO-BP}	Fitness _{ABC-BP}	Fitness _{IABC-BP}
1	0.6770	0.7678	0.9103	0.9606
2	0.9372	0.9515	0.9271	0.9611
3	0.9443	0.9515	0.9272	0.9644
4	0.9461	0.9533	0.9341	0.9667
5	0.9469	0.9537	0.9341	0.9667
10	0.9493	0.9554	0.9501	0.9678
20	0.9507	0.9572	0.9578	0.9681
50	0.9514	0.9574	0.9616	0.9690
100	0.9515	0.9576	0.9632	0.9693
200	0.9515	0.9577	0.9648	0.9702
300	0.9515	0.9577	0.9652	0.9704
400	0.9515	0.9577	0.9656	0.9705
500	0.9515	0.9577	0.9657	0.9705

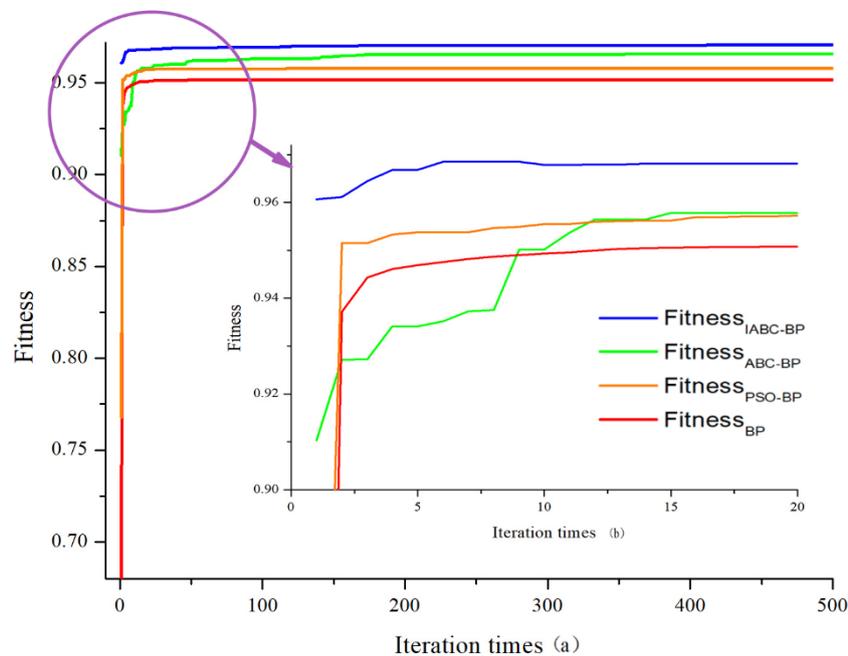


Figure 6. The comparison of fitness among the four models. (a) The convergence performance from the first iteration time to the last of four models; (b) The convergence performance from the first iteration time to the twentieth iteration time.

4.1.1. Convergence Accuracy

The higher the convergence accuracy, the more accurate is the model. In this study, fitness represents the convergence accuracy, the closer the fitness is to 1, the higher the accuracy the model is equipped with. According to Table 3 and Figure 6, the accuracies of BP, PSO-BP, ABC-BP, and IABC-BP are 0.9515, 0.9577, 0.9657, and 0.9705 respectively. Fitness (IABC-BP) gets a much more accurate value at the first iteration time and holds on to its favorable position all along.

4.1.2. Convergence Speed

The fewer iteration times the model needs to converge to a stable value, the faster is the convergence speed of the model. As shown in Table 3 and Figure 6, Fitness (BP) and Fitness (PSO-BP) converge quickly. Compared with Fitness (ABC-BP), Fitness (IABC-BP) converges much faster, it approaches the stable value 0.97 around the 200th iteration time, however Fitness (ABC-BP) tends to be stable until the 400th iteration time.

In conclusion, from the perspective of convergence performance, IABC-BP algorithm model is superior to the others.

4.2. Results Analysis

Four models of BP, PSO-BP, ABC-BP, and IABC-BP neural network were constructed for the data of the factors affecting the water quality. The first 50 groups of the 66 data were used to train the network of the four models, and the latter 16 groups were used to test the network. The compared results of each model and each factor are shown in Figures 7–9. According to the comparison, it is obvious that the blue lines which represent the predicted data by IABC-BP neural network model are closer to the black lines that represent the actual data. The red lines which represent predicted data by BP neural network deviate a lot from the black ones. Therefore, a rough conclusion can be drawn that the predicting capability of the IABC-BP neural network model is the most appropriate of these four models and the BP is relatively worse than the others.

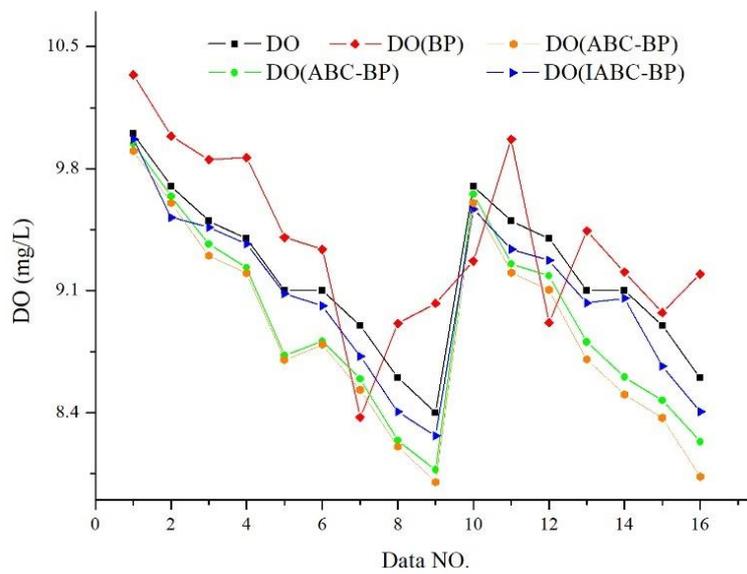


Figure 7. Comparison of DO among the four models.

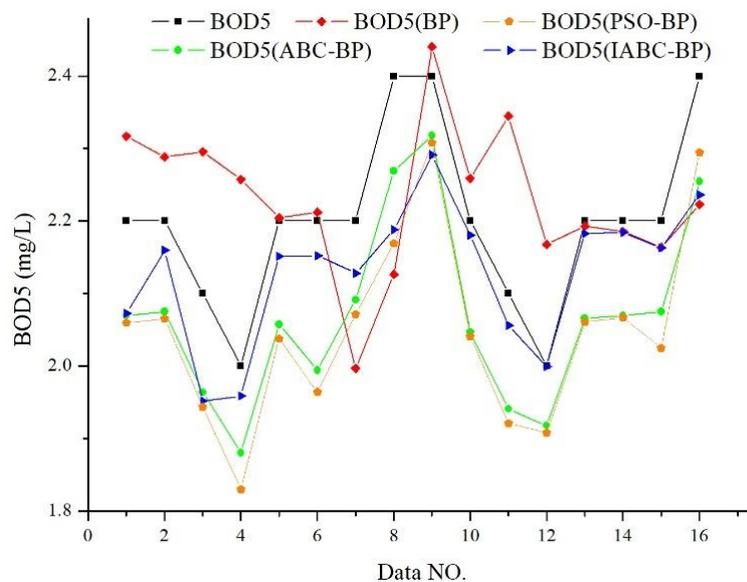


Figure 8. Comparison of BOD₅ among the four models.

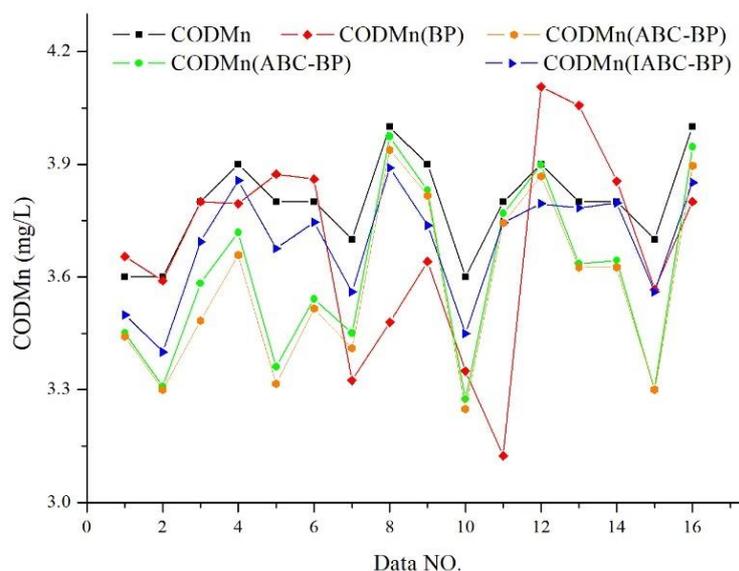


Figure 9. Comparison of COD_{Mn} among the four models.

In order to illustrate the conclusion more clearly, the results of DO were selected for further analysis from the following three aspects.

4.2.1. Relative Error (RE)

The separate comparison between actual DO values and the four models are shown in Figures 10–13. Through the analysis of the absolute relative error value of the BP neural network model, we found that there were three absolute values of the relative prediction error less than 3%, seven values between 3% and 5%, and six values above 5%. Based on the analysis of the PSO-BP network model, we found that there were five absolute values of the relative prediction error less than 3%, nine values between 3% and 5%, and only three above 5%; Based on the analysis of the ABC-BP network model, we found that there were seven absolute values of the relative prediction error less than 3%, eight values between 3% and 5%, and only one above 5%. In the IABC-BP model, there were 14 absolute values of the relative prediction error less than 3%, two between 3% and 5%, and none above 5%. It shows that the error of each index in the IABC-BP neural network model is almost the smallest of the four models, which indicates that the IABC-BP model is the most suitable model of the four models. The BP model of predicting water quality is comparatively the worst, although its convergence speed is high and the predicted results can be obtained quickly. For the PSO-BP model and the ABC-BP model, in most cases, the predicted results are acceptable. By comparing the BP model with ABC-BP model, the prediction effect of ABC-BP is better than that of BP. Therefore, the process of finding the proper connection weights and thresholds in BP requires the ABC algorithm. By comparing the PSO-BP model with the ABC-BP model, the prediction effect of ABC-BP is better than that of PSO-BP, while the convergence speed of PSO-BP is much faster than ABC-BP in the beginning. Therefore, combined with the inter-cell generation method and inspired by the PSO algorithm, the prediction effect of the IABC-BP model approaches a more accurate result, indicating that IABC has greatly improved ABC, and it is certainly applicable to improve the BP neural network prediction model.

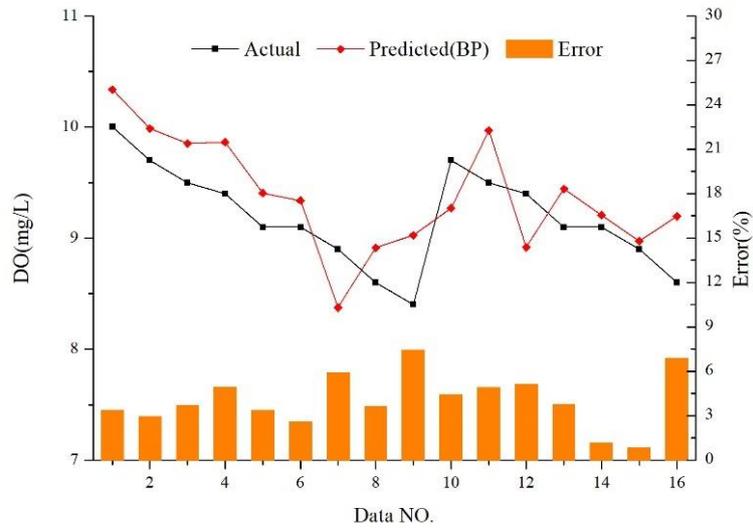


Figure 10. The predicted BP value and its error percentage.

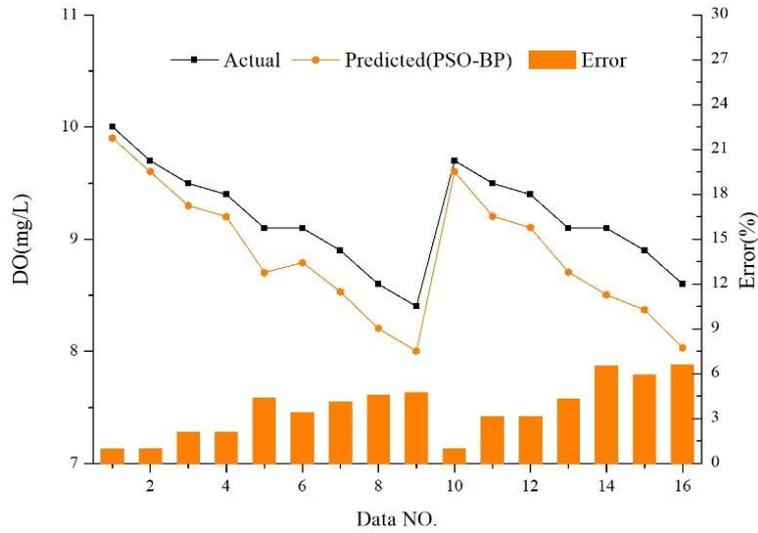


Figure 11. The predicted PSO-BP value and its error percentage.

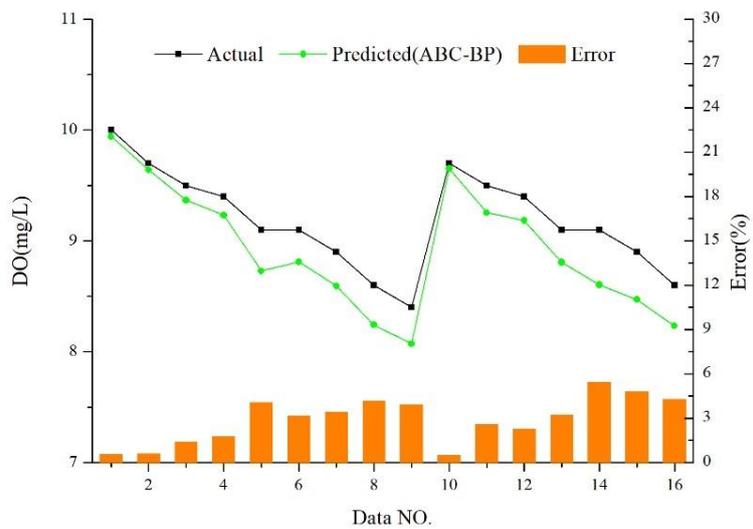


Figure 12. The predicted ABC-BP value and its error percentage.

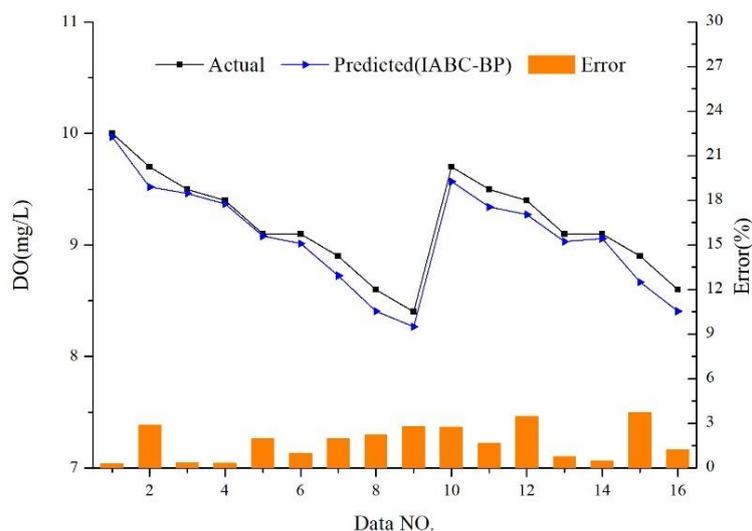


Figure 13. The predicted IABC-BP value and its error percentage.

4.2.2. Coefficient of Determination R^2

In addition to relative errors, the R^2 values were also computed according to the results in order to access the accuracy of the model. The closer R^2 is to 1, the higher is the reference value of the relevant equation. Conversely, the closer R^2 gets to 0, the lower the reference value. As shown in Table 4, the corresponding values of BP, PSO-BP, ABC-BP, and IABC-BP were 0.658, 0.918, 0.942, and 0.981 which shows the IABC-BP predicting model fits better than the other three models.

Table 4. Coefficient of determination R^2 of each model.

Name of Model	BP	PSO-BP	ABC-BP	IABC-BP
R^2	0.658	0.918	0.942	0.981

4.2.3. Nash–Sutcliffe Efficiency Coefficient (NSE)

The NSE coefficient of the BP, PSO-BP, ABC-BP, and IABC-BP models were 0.134, 0.296, 0.541, and 0.805 as shown in Table 5. The results illustrate that the predicting results of all these four models are credible and reliable. Besides, the IABC-BP model is the model equipped with the best quality and the highest reliability, followed by ABC-BP, PSO-BP, and BP in turn.

Table 5. NSE coefficient of each model.

Name of Model	BP	PSO-BP	ABC-BP	IABC-BP
NSE	0.134	0.296	0.541	0.805

To conclude, from the perspective of the three aspects mentioned above, the IABC-BP model has been improved substantially based on the ABC-BP model and shows its superiority in both performance of convergence and results.

We can see in Figures 7–9 that the forecasting performance of DO is the optimum training result while the performance of BOD and COD_{Mn} sometimes would seriously deviate from the actual lines. The BP neural network would become very unstable if the training samples were not enough, because it requires a large number of data for training. However, the ER project has not been completed for a long enough time and the monitoring system of the water quality has not been comprehensive enough to provide sufficient data, so that the amount of available data is limited, thus the forecast would sometimes underperform. Therefore, a more accurate model which has the ability to map nonlinear

data, functions better in situations of data deficiency and also is capable of coping with unstable data needs to be further explored.

5. Conclusions

This study focuses on whether an IABC-BP model can be used to forecast water quality under normal conditions in the ER project which has stable water quality due to its design and whose water quality can be measured continuously as the indices that influence water quality are measurable. The data which were acquired from the east route were used to test and verify our hypothesis and the conclusions were as follows.

Compared with the ABC-BP model, the IABC-BP model which can increase the forecasting performance of the ABC-BP by searching for the best value of each connection weight and threshold has better network stability, higher learning speed, and stronger approximation ability. In a word, the IABC-BP model has been improved effectively, and it is superior to ABC-BP in all aspects and more suitable for water quality prediction.

To sum up, the IABC-BP model is effective for predicting water quality accurately and rapidly under normal conditions. In this study, the feasibility of water quality prediction was verified through systematic method, and a new concept of water quality prediction provided for practical application. The prediction model is applicable for the ER project of the SNWD project, and it can obtain water quality of the sixth-staged pumping station to ulteriorly predict the water quality of the intake area and improve the water quality which may not be meeting the standards in a specific way. Therefore, this paper plays a significant role in the water security for the East Route of the South-to-North Water Diversion Project.

Author Contributions: Conceptualization, S.C. and G.F.; Methodology, S.C.; Software, S.C.; Validation, X.H., Y.Z. and G.F.; Formal Analysis, S.C.; Investigation, Y.Z.; Resources, S.C.; Data Curation, X.H.; Writing-Original Draft Preparation, S.C.; Writing-Review & Editing, G.F.; Visualization, S.C. and Y.Z.; Supervision, G.F.; Project Administration, G.F.; Funding Acquisition, G.F.

Funding: This research was funded by Jiangsu Water Conservancy Science and Technology Project (NO. 2014012).

Acknowledgments: Thanks are due to Chen, H. for assistance with the use of software, to You, C.X. for providing the experimental data and to the referees and the editor for their valuable comments and suggestions on improvement of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Li, Y.S.; Liu, W.J.; Wu, Y.Y.; Wang, D. A review of progress in research on water quality models in America. *Water Resour. Hydropower Eng.* **2006**, *2*, 15. (In Chinese)
- Xiang, N. Exploration for Water Quality Assessment and Prediction based on Neural Networks and Artificial Bee Colony Algorithm. Ph.D. Thesis, South China University of Technology, Guangzhou, China.
- Qiu, L.; Huang, X.; Li, H.L. Comprehensive water quality prediction based on fuzzy weight Markov model. *Yangtze River* **2007**, *38*, 75–77. (In Chinese)
- Lettenmaier, D.P. Detection of trends in water quality data from records with dependent observations. *Water Resour. Res.* **1976**, *12*, 1037–1046. [[CrossRef](#)]
- Liang, J. Application of Support Vector Machine in Water Quality Evaluation and Prediction. Ph.D. Thesis, Zhejiang University of Technology, Hangzhou, China, 2009.
- Ju, Z.C.; Wang, X.; Gong, Y.X. Prediction of water quality in Yellow River based on BP neural network model. *J. Qinghai Univ.* **2017**, *35*, 88–92.
- Yang, Y.M. Researches on Extreme Learning Theory for System Identification and Applications. Ph.D. Thesis, Hunan University, Changsha, China, 2013.
- Maier, H.R.; Dandy, G.C. The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* **1996**, *32*, 1013–1022. [[CrossRef](#)]
- Cao, J.H.; Lin, H.W.; Zhang, S.C. Prediction of water quality index in Danjiangkou reservoir based on BP neural network. *Electron. Des. Eng.* **2010**, *18*, 17–19. (In Chinese)

10. Chang, Q.C. BP neural network improved algorithm to predict the application of water quality. *Gansu Environ. Study Monit.* **2002**, *3*, 186–188. (In Chinese)
11. Chen, D.Y.; Zhang, X.Z. Application of variable structure neural network in prediction of future water quality parameters. *Sci. Technol. Eng.* **2008**, 1577–1579. (In Chinese)
12. Singh, K.P.; Basant, A.; Malik, A.; Jain, G. Artificial neural network modeling of the river water quality—A case study. *Ecol. Model.* **2009**, *220*, 888–895. [[CrossRef](#)]
13. Zheng, G.Y.; Luo, F.; Chen, W.B. Quality prediction of waste water treatment based on Immune Particle Swarm Neural Networks. *Microprocessors* **2010**, *31*, 75–77. (In Chinese)
14. Gao, F.; Feng, M.Q.; Teng, S.F. On the way for forecasting the water quality by BP neural network based on the PSO. *J. Saf. Environ.* **2015**, *15*, 338–341.
15. Zhang, X.D.; Gao, M.T. Water quality prediction method based on IGA-BP. *Chin. J. Environ. Eng.* **2016**, *10*, 1566–1571.
16. Wang, Z.; Shao, D.; Yang, H.; Yang, S. Prediction of water quality in South to North Water Transfer Project of China based on GA-optimized general regression neural network. *Water Sci. Technol. Water Supply* **2015**, *15*, 150–157. [[CrossRef](#)]
17. Huo, J.; Liu, L.; Zhang, Y. An improved multi-cores parallel artificial Bee colony optimization algorithm for parameters calibration of hydrological model. *Future Gener. Comput. Syst.* **2018**, *81*, 492–504. [[CrossRef](#)]
18. Karaboga, D.; Basturk, B. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *J. Glob. Optim.* **2007**, *39*, 459–471. [[CrossRef](#)]
19. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput.* **2008**, *8*, 687–697. [[CrossRef](#)]
20. Karaboga, D.; Akay, B. A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **2009**, *214*, 108–132. [[CrossRef](#)]
21. Su, C.H.; Xiang, N.; Chen, G.Y.; Wang, F. Water quality evaluation model based on artificial bee colony algorithm and BP neural network. *Chin. J. Environ. Eng.* **2012**, *6*, 699–704.
22. Zhang, Z.S.; Hu, Y.G.; Zhao, H.W.; Feng, Y.Z.; Feng, P.Q. A Study on the Improved Genetic Algorithm and Its Implementation. *Microelectronics* **2002**, *32*, 273–275.
23. Bao, L.; Zeng, J.C. A bi-group differential artificial bee colony algorithm. *Control Theory Appl.* **2011**, *28*, 266–272.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).