A Fusion Water Quality Soft-Sensing Method Based on WASP Model and Its Application in Water Eutrophication Evaluation

Xiaoyi Wang, Jie Jia, Tingli Su, Zhiyao Zhao, Jiping Xu, and Li Wang

School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

Correspondence should be addressed to Zhiyao Zhao; zhaozy@bitbu.edu.cn

Received 7 November 2017; Revised 24 January 2018; Accepted 6 February 2018; Published 6 March 2018

Copyright © 2018 Xiaoyi Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Water environment protection is of great significance for both economic development and improvement of people's livelihood, where modeling of water environment evolution is indispensable in water quality analysis. However, many water quality indexes related to water quality model cannot be measured online, and some model parameters always vary among different water areas. Thus, this paper proposes a water quality soft-sensing method based on the water quality mechanism model to simulate evolution of water quality indexes online, where unscented Kalman filter is utilized to estimate model parameters. Furthermore, a modified fuzzy comprehensive evaluation method is presented to evaluate the level of water eutrophication condition. Finally, the water quality data collected from Taihu Lake and Beihai Lake are used to validate the effectiveness and generality of the proposed method. The results show that the proposed soft-sensing method is able to describe the variation of related water quality indexes, with better accuracy compared to nonlinear least squares based method and traditional trial-and-error based method. On this basis, the water eutrophication condition can be also accurately evaluated.

1. Introduction

With the rapid development of modern society, production of industrial and sanitary sewage is daily increasing, and eutrophication phenomenon of lakes and reservoirs is becoming much more serious [1, 2]. Generally, the occurrence of eutrophication is related to excessive nitrogen, phosphorus, and other inorganic nutrients in water, where nitrogen and phosphorus are the main reasons accounting for the eutrophication of slow flow water, such as lakes, reservoirs, and bays [3–5]. Currently, eutrophication phenomenon exists in 54% of the lakes of Asia-Pacific region [6]. Therefore, how to economically and effectively handle the eutrophication problem has become an urgent priority.

In order to timely evaluate or predict the eutrophication condition, the variation of water quality indexes should be timely measured or learned [7, 8]. However, many water quality indexes, such as biochemical oxygen demand (BOD) and total nitrogen (TN), cannot be measured online [9]. Thus, the manner of soft-sensing is introduced to overcome this limitation in this paper. Soft-sensing is to establish a mathematical relation model between easily measured process variables and difficultly measured process variables based on mechanism analysis and sensor data mining [10]. The existing soft-sensing modeling approach can be classified into three types: mechanism modeling, identification modeling, and artificial intelligence-based modeling [11–13]. The mechanism modeling approach is to obtain a mathematical expression based on the analysis of the system's internal relations, which adopts the basic physical and chemical laws, such as material, energy, or momentum conservation relation [14]. The identification modeling approach is to establish a mathematical model based on the information of system input and output by certain parameter identification, filtering, or regression analysis methods, without understanding the mechanism of the dynamic process [15]. The artificial intelligence-based modeling approach is to get an underlying model of the real-world system or a portion of system based on artificial intelligent methods [16]. In addition, there also exist fusion methods derived from a combination of the above approaches; that is, the mechanism modeling approach is utilized to describe partial behavior of the studied
system with known mechanism, and identification modeling approach or artificial intelligence-based approach is used to handle the remaining part.

Water Quality Analysis Simulation Program (WASP) mechanism model is a comprehensive water quality model that can be used to interpret the process of natural or artificial water quality deterioration [17]. It can simulate migration and transformation of conventional water quality indexes (including dissolved oxygen (DO), BOD, and nutrients) and toxic contaminants (including organic chemicals, metals, and sediment) in water [18]. Currently, WASP model has been widely applied to different water areas, such as Mobile Bay [19], Murderkill River [20], Lake Michigan [21], and Songhua River [22]. However, WASP also has a limitation. That is, some model parameters vary among different water areas, and their values are always determined by trial-and-error method [23]. This is insufficient for accurately modeling the water quality variation of a specific water area.

Therefore, this paper builds a fusion water quality soft-sensing method, where the WASP model is employed as a soft-sensing method and its unknown parameters are estimated by the unscented Kalman filter (UKF) [24]. Then, the variations of DO, BOD, nitrate nitrogen (NO$_3$-N, related to TN), ammonia nitrogen (NH$_3$-N), phytoplankton carbon (Phyt, related to chlorophyll$_a$ (Chl$_a$)), and so forth can be simulated by the fusion water quality soft-sensing method. On this basis, a modified fuzzy comprehensive evaluation method is presented to evaluate the eutrophication condition of the rivers and lakes, combining both the simulated values and the online measured values of transparency (SD) and total phosphorus (TP) [25]. Finally, by taking Taihu Lake and Beihai Lake as examples, the effectiveness and generality of the proposed method are validated and the water eutrophication condition is evaluated. Comparative studies are also presented and discussed.

The remainder of this paper is organized as follows. Section 2 presents the methodology of the WASP based water quality soft-sensing method, where a simplified WASP model is presented and the procedure of unknown parameters estimation by UKF is listed. Section 3 introduces the modified fuzzy comprehensive evaluation method, where the modified methods of selecting water quality indexes and calculation of the corresponding weight are presented. Section 4 presents two case studies of Taihu Lake and Beihai Lake to validate the effectiveness and generality of the proposed fusion soft-sensing method and the water eutrophication evaluation method. Section 5 gives the conclusion and indicates future development.

2. Fusion Water Quality Soft-Sensing Method

2.1. Dynamic Model of Water Quality Based on WASP

Eutrophication module (shorten as EUTRO) is an essential part of the WASP model. This module describes the dynamic behavior of water quality indexes including DO, BOD, Phyt, NO$_3$-N, NH$_3$-N, organic nitrogen (ON), and organic phosphorus (OP). The interacting relations of them can be represented by four reaction systems, namely, phytoplankton kinetics, phosphorus cycle, nitrogen cycle, and DO balance.

In the EUTRO module, a simplified dynamic process of seven water quality indexes are presented as [18]

\[ x = \begin{pmatrix} C_{DO} \\ C_{BOD} \\ C_{Phyt} \\ C_{NO3-N} \\ C_{NH3-N} \\ C_{ON} \\ C_{OP} \end{pmatrix}, \]

where \( x = (C_{DO}, C_{BOD}, C_{Phyt}, C_{NO3-N}, C_{NH3-N}, C_{ON}, C_{OP})^T \) represent concentrations of DO, BOD, Phyt, NO$_3$-N, NH$_3$-N, ON, and OP, respectively. In (1), the chemical meaning of unknown model parameters \( \{k_1, k_2, \ldots, k_{10}\} \) is shown in Notations [18].

Let \( \theta = (k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10})^T \). Here, assume that the unknown parameters are piecewise-constant. Then, the process equation can be written as

\[
\begin{bmatrix}
\dot{x} \\
\dot{\theta}
\end{bmatrix} = 
\begin{bmatrix}
f(x) \\
0_{10 	imes 1}
\end{bmatrix},
\]

Further let \( X = [x^T, \theta^T]^T \) and the process noise item be added. An augmented process equation is obtained as [26]

\[
X = 
\begin{bmatrix}
f(x) \\
0_{10 	imes 1}
\end{bmatrix} + w,
\]

where \( w \) is the process noise, satisfying that \( w \sim N(0, Q) \). \( Q \) is the covariance matrix. Then, the observation equation is set as follows:

\[
y = [h \\ 0] x + v,
\]

where \( h \) is the observation matrix, describing the mapping from state indexes to observations; \( v \) is the measurement noise, satisfying that \( v \sim N(0, R) \). \( R \) is the covariance matrix.
Up to now, we obtain a continuous-time dynamic model of water quality indexes as follows:

\[ \dot{X} = \begin{bmatrix} f(x) \\ 0_{10 \times 1} \end{bmatrix} + u, \]
\[ y = [h \ 0] \begin{bmatrix} x \\ \theta \end{bmatrix} + v. \]  

(5)

To obtain the discrete-time model, the fourth-order Runge-Kutta is utilized, where the step length is \( l \). The formula is as follows:

\[
\begin{align*}
K_1 &= f (X_k), \\
K_2 &= f \left( X_k + \frac{l}{2} K_1 \right), \\
K_3 &= f \left( X_k + \frac{l}{2} K_2 \right), \\
K_4 &= f \left( X_k + l K_3 \right), \\
X_{k+1} &= X_k + \frac{l}{6} (K_1 + 2K_2 + 2K_3 + K_4).
\end{align*}
\]  

(6)

Then, (5) is discretized as

\[
\begin{align*}
X(k) &= f (X(k-1)) + w(k-1), \\
y(k) &= [h \ 0] X(k) + v(k).
\end{align*}
\]  

(7)

2.2. Unknown Parameter Estimation Based on UKF. In practice, UKF is an effective tool for state estimation [27]. Given the dynamic equation shown in (7) and a series of observations \( Y(k) = [y(1), y(2), \ldots, y(k)] \), the state estimation procedure by the UKF is presented as follows [28].

Step 1 (UT transformation). In the original state distribution, some sampling points are selected according to certain rules, so that the mean and covariance of these sampling points are equal to the mean and covariance of the previous state distribution. These points are substituted into the nonlinear function, and the corresponding set of the nonlinear function values is obtained. Then, the mean and covariance of the nonlinear transformation are obtained from these sets of points.

First, compute the \( 2n + 1 \) sigma points, namely, sampling points:

\[
\begin{align*}
X^{(0)}_{k-1} &= X^a_{k-1}, \\
X^{(i)}_{k-1} &= X^a_{k-1} + \left( \sqrt{(n+\lambda) P_{k-1}} \right)_i, \\
X^{(m-i)}_{k-1} &= X^a_{k-1} - \left( \sqrt{(n+\lambda) P_{k-1}} \right)_i, \quad i = 1, \ldots, n,
\end{align*}
\]  

(8)

where \( n \) is the number of the state dimension.

Second, calculate the associated weight \( w \) of the sampling points:

\[
\begin{align*}
w^{(0)}_m &= \frac{\lambda}{n+\lambda} \\
w^{(i)}_c &= \frac{\lambda}{n+\lambda} + (1-\alpha^2 + \beta) \\
w^{(i)}_w &= \frac{1}{2(n+\lambda)} \quad i = 1, \ldots, 2n \\
\lambda &= \alpha^2 (n + \kappa) - n,
\end{align*}
\]  

(9)

where \( P \) is the estimated covariance, satisfying \((\sqrt{P})^T (\sqrt{P}) = P\); \( \sqrt{P} \) is the square root of the \( i \)th column in \( P \); \( \lambda \) is scaling parameter used to reduce the total prediction error; \( \alpha \) controls the spread of sampling points; \( \kappa \) is the selected parameter, and its value is not bounded generally, but it is usually necessary to ensure the semidefinite of \((n + \lambda)P\) matrix. Under normal circumstances, \( \alpha = 10^{-3} \) and \( \kappa = 0 \). \( \beta \) is a nonnegative weighting index, and the optimal value is 2 for a Gaussian distribution of \( X \).

Step 2. Compute predicted state \( \hat{X}^{(i)}_k \) and predicted covariance \( P^{(i)}_k \):

\[
\begin{align*}
\hat{X}^{(i)}_k &= f (\hat{X}^{(i)}_{k-1}), \\
X^f_k &= \sum_{i=0}^{2n} w^{(i)} \hat{X}^{(i)}_k, \\
P^f_k &= \sum_{i=0}^{2n} w^{(i)} \left( \hat{X}^{(i)}_k - X^f_k \right) \left( \hat{X}^{(i)}_k - X^f_k \right)^T.
\end{align*}
\]  

(10)

Step 3. Compute predicted \( \hat{y}_k \), measurement covariance \( P_{\text{yy}_k} \), and cross-covariance of the state and measurement \( P_{\text{xy}_k} \):

\[
\begin{align*}
\hat{\gamma}^{(i)}_k &= \hat{X}^{(i)}_k, \\
\hat{\gamma}_k &= \sum_{i=0}^{2n} w^{(i)} \hat{\gamma}^{(i)}_k, \\
P_{\gamma y_k} &= \sum_{i=0}^{2n} w^{(i)} \left( \hat{\gamma}^{(i)}_k - \hat{\gamma}_k \right) \left( \hat{\gamma}^{(i)}_k - \hat{\gamma}_k \right)^T, \\
P_{\gamma x_k} &= \sum_{i=0}^{2n} w^{(i)} \left( \hat{\gamma}^{(i)}_k - X^f_k \right) \left( \hat{\gamma}^{(i)}_k - X^f_k \right)^T.
\end{align*}
\]  

(11)

Step 4. Compute gain \( K_k \), updated state \( X^u_k \), and covariance \( P_k \):

\[
\begin{align*}
K_k &= P_{\gamma x_k} P_{\gamma y_k}^{-1}, \\
X^u_k &= X^f_k + K_k (y_k - \hat{y}_k), \\
P_k &= P^f_k - K_k P_{\gamma y_k} K_k^T.
\end{align*}
\]  

(12)
From the estimate of the augmented variable $X$, the estimated values of unknown parameters can be obtained as
$$\hat{\theta} = [\hat{k}_1, \hat{k}_2, \ldots, \hat{k}_M].$$

### 3. Eutrophication Condition Evaluation

#### Based on Modified Fuzzy Comprehensive Evaluation Method

On the basis of the output from the proposed fusion water quality soft-sensing method and online measurements of water quality indexes, a modified fuzzy comprehensive evaluation method is used to evaluate the water eutrophication condition.

Fuzzy comprehensive evaluation method adopts fuzzy mathematical theory to obtain a quantitative evaluation result of an object in view of the complexity of object and fuzziness of the evaluation index. The procedure of the proposed fuzzy comprehensive evaluation method is presented as follows.

**Step 1 (water evaluation index selection).** In the traditional fuzzy comprehensive evaluation method, the key evaluation indexes, which have great influence on the water environment, are obtained by empirical measures. Although these methods are relatively convenient, they lack objectivity and theoretical foundation. In order to compensate this limitation, a cumulative frequency method is introduced by calculating the cumulative frequency of excessive multiple of each water quality index [29].

The evaluation set is a collection of criteria for evaluating the object [30]. Suppose the water eutrophication condition can be classified to $n$ levels, written as
$$V = \{V_1, V_2, \ldots, V_n\}.$$  \hspace{1cm} (13)

Further suppose there exist $M$ evaluation indexes, written as $\{C_1, C_2, \ldots, C_M\}$. On this basis, the cumulative frequency can be calculated as follows:

$$\beta_i = \frac{C_i}{(1/n) \sum_{j=1}^{n} \sigma_{ij}},$$

$$\bar{\beta}_i = \{\beta_1 \mid \beta_i \geq \beta_i \geq \beta_M\}$$

$$i = 1, 2, \ldots, M; \ j = 1, 2, \ldots, n,$$  \hspace{1cm} (14)

$$K_j = \frac{\sum_{i=1}^{J} \bar{\beta}_j}{\sum_{M} \bar{\beta}_j} \times 100\%,$$

where $i$ is the label of the evaluation index; $j$ is the label of eutrophication level; $C_i$ is the concentration value of the $i$th index; $\sigma_{ij}$ is the standard value of the $i$th index in level $j$; $\beta_i$ is the excessive multiple value of the $i$th index; $K_j$ is the cumulative frequency of the first $i$ indexes. According to the statistical analysis requirements in the selection of evaluation indexes, generally take [31]

$$K_j \geq 85\%.$$  \hspace{1cm} (15)

Following (14) and (15), we have the selected key water evaluation indexes as $C = \{C_1, C_2, \ldots, C_m\}$, $m \leq M$.

**Step 2** (establishment of the fuzzy relation matrix $R : C \rightarrow V$). By adopting fuzzy mathematical theory to evaluation study, the most critical issue is to establish the membership functions for the evaluation indexes. Triangular linear membership functions are commonly used in practice, which are also selected for determining the fuzzy relation matrix $R$ in this paper [32]. The configuration of the membership function is as follows:

For $\forall V_j \in V (j = 1, 2, \ldots, n)$,

(1) when $j = 1$, the membership function is

$$r_{ij} = \begin{cases} 1 & 0 \leq C_i \leq \sigma_{ij} \\ \frac{\sigma_{ij} - C_i}{\sigma_{ij} - \sigma_{ij+1}} & C_i > \sigma_{ij+1}, \end{cases} \hspace{1cm} (16)$$

(2) when $1 < j < n - 1$, the membership function is

$$r_{ij} = \begin{cases} 1 & C_i = \sigma_{ij} \\ \frac{C_i - \sigma_{ij-1}}{\sigma_{ij} - \sigma_{ij-1}} & \sigma_{ij-1} \leq C_i < \sigma_{ij} \\ \frac{C_i - \sigma_{ij+1}}{\sigma_{ij+1} - \sigma_{ij}} & \sigma_{ij} < C_i \leq \sigma_{ij+1} \\ 0 & C_i < \sigma_{ij-1}, C_i > \sigma_{ij+1}, \end{cases} \hspace{1cm} (17)$$

(3) when $j = n$, the membership function is

$$r_{ij} = \begin{cases} 1 & C_i \geq \sigma_{ij} \\ \frac{C_i - \sigma_{ij-1}}{\sigma_{ij} - \sigma_{ij-1}} & \sigma_{ij-1} \leq C_i < \sigma_{ij} \\ 0 & C_i < \sigma_{ij-1}, \end{cases} \hspace{1cm} (18)$$

where $r_{ij}$ presents the membership degree of index $C_i$ in level $V_j$. $\sigma_{ij-1}, \sigma_{ij}, \sigma_{ij+1}$ are the standard values of the $i$th index in level $V_{j-1}, V_j, V_{j+1}$, respectively. When $C_i$ is given, the above membership functions can be applied to determine the membership degree of the evaluation index $C_i$ for each level of water eutrophication.

Then, the fuzzy relation matrix $R : C \rightarrow V$ is constructed as

$${R} = (r_{ij})_{m \times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}, \hspace{1cm} (19)$$

where $\sum_{j=1}^{n} r_{ij} = 1, \ i = 1, 2, \ldots, m$.

**Step 3** (weight determination of evaluation index). The determination of the evaluation index weight is one of the most important factors that directly affect the final evaluation results. In this paper, the clustering weight method is used to determine the weight of each evaluation index, combining
The index concentration value with the standard values which more objectively reflect the relative importance of each evaluation index in all the indexes. The method of calculating the index weight is as follows:

\[ w_{ij} = \frac{C_i/\sigma_{ij}}{\sum_{i=1}^{m} C_i/\sigma_{ij}} \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n \]  \hspace{1cm} (20)

where \( w_{ij} \) is the weight of the \( i \)th water evaluation index in the \( j \)th eutrophication level. Therefore, the index weight matrix \( W \) determined by the clustering method is

\[ W = (W_1, W_2, \ldots, W_j, \ldots, W_n), \]  \hspace{1cm} (21)

where \( W_j \) is the index weight matrix of the \( j \)th eutrophication level.

**Step 4** (fuzzy synthesis operation). Combining the index weight \( W \) with the fuzzy relation matrix \( R \), the multiplication and addition method of the weighted average is chosen to obtain fuzzy comprehensive evaluation result \( B \) based on all indexes. The advantage of this method is that it can balance all indexes according to the weight values to reflect the comprehensive condition of water quality [33]. The specific formula is as follows:

\[ B = \left(b_j\right)_{1 \times n} = W \circ R, \]

\[ b_j = \sum_{i=1}^{m} w_{ij} \cdot r_{ij}, \]  \hspace{1cm} (22)

where the element \( b_j \) is membership of the water object with regard to \( j \)th water eutrophication level. The water eutrophication level can be obtained by the principle of maximum membership, where the specific formula is as follows:

\[ j = \arg \max_{j} b_{j}. \]  \hspace{1cm} (23)

### 4. Case Study

In this part, water quality data of Taihu Lake (case 1) and Beihai Lake (case 2) are utilized to verify the effectiveness and generality of the proposed method. First the UKF is used to estimate the unknown parameters in the soft-sensing method, where the estimated result is compared to the results obtained by the nonlinear least squares method and trial-and-error method. Then, simulated values of the water quality indexes are deduced, which are compared to the real measured water quality data. On this basis, the water eutrophication evaluation is carried out by the modified fuzzy comprehensive evaluation method, depending on both simulated values of DO, BOD, TN, Chl_a and real measured values of SD, TP.

#### 4.1. Case 1: Taihu Lake

1. **Model Parameter Estimation Result and Analysis.** The UKF is used to estimate the unknown model parameters of Taihu Lake shown in Notations. Besides, due to the limitation of measured data, the TN data are used instead of NO_3-N. Since the measured data include DO, BOD, Phyt, TN, and NH_3-N, the observation matrix in (7) is set as

\[ h = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{5 \times 17} \]  \hspace{1cm} (24)

Then the real-time estimated values of the ten unknown model parameters can be obtained, which are depicted in Figure 1.

The appearance of algal blooms is a feature of water eutrophication, whose formation process can be divided into three stages, namely, recovery, biomass increase and accumulation, dormancy [34]. In this paper, January to March is the recovery period of algae bloom, April to mid-October is the second stage, and the remaining part is the dormancy stage. Following the multistage principle, the average values of unknown model parameters of each stage are shown in Table 1.

Then, the estimated values of the unknown parameters in Table 1 are substituted into the water quality soft-sensing method, and the simulation process is carried out. For comparison, the simulated values of water quality indexes are compared to the real measured values and the simulated values obtained by the models with estimated parameters based on the trial-and-error method and nonlinear least squares method, respectively. Figure 2 depicts the results of DO concentration, BOD concentration, NH_3-N concentration, TN concentration, and Chl_a concentration. It should be noted that the amount of Chl_a is indirectly expressed as the concentration of Phyt; that is to say, through the ratio between Phyt and Chl_a, the concentration of Chl_a is obtained.

It can be drawn from Figure 2 that the simulated values are in good agreement with the measured values of water
quality indexes, besides TN concentration. There are two main reasons for this. One is that the data of TN are used instead of NO$_3$-N during the experiment, and the other is the inaccuracies caused by external influences. However, the overall experimental results can verify the effectiveness of the fusion water quality soft-sensing method. Then the model accuracy based on the UKF is better than those based on the nonlinear least squares method and trial-and-error method. In order to quantitatively evaluate the error, Root Mean Square Error (RMSE) is utilized to indicate the deviation between the simulated values and the measured values of each water quality index. The specific formula is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \hat{X}_i)^2}{n}},$$

where $X_i$ is the measured value, $\hat{X}_i$ is the simulated value, and $n$ is the number of measurement. The result is shown in Table 2.

**4.1.2. Water Eutrophication Evaluation Result and Analysis.**

The modified fuzzy comprehensive evaluation method is used to evaluate the water eutrophication status of Taihu Lake, by taking both the simulated values and measured values of the water quality indexes into consideration. For comparison, the eutrophication evaluation result is compared to the result based on merely measured data.

**Step 1** (water evaluation index selection). As the eutrophication mechanism is complex, scholars often choose different evaluation criteria to assess water quality status. By referring to [35, 36], five levels are classified to describe the eutrophication condition: I (none), II (mild), III (medium), IV (heavy), and V (extremely heavy); namely, $V = \{I, II, III, IV, V\}$.

Then, according to (14) and (15) to select the water quality evaluation indexes, SD, BOD, TN, TP, DO, and Chl$_a$ are selected as evaluation indexes. By referring to the technological regulations for surface water resources quality assessment published by Ministry of Water Resources, People's Republic of China, and related references [37, 38], the standard values of water quality indexes for lakes and reservoirs in each eutrophication level can be determined, which are shown in Table 3. On this basis, triangular linear membership functions of the water quality indexes are constructed as shown in Figure 3.

**Step 2** (determine the fuzzy relation matrix $R(t)$ and the weight matrix $W$). Given the data collected from a monitoring station in Taihu Lake, according to (16)–(18) and (20), the time-related fuzzy relation matrix $R(t)$ and weight matrix $W(t)$ are obtained, respectively.
Step 3 (fuzzy synthesis operation). The fuzzy comprehensive evaluation result $B(t)$ of the monitoring station in Taihu Lake is obtained by using (22). Figure 4 depicts the obtained memberships of level I, level II, level III, level IV, and level V based on the measured values and simulated values of UKF, respectively.

According to (23), the water eutrophication level based on measured values, simulated values of trial-and-error method, simulated values of nonlinear least squares, and simulated values of UKF is evaluated, and results are shown in Figure 5.

It can be seen from Figure 5 that, during the recovery and dormancy stage of algal blooms, the water eutrophication status is basically in level I, while in the biomass increase and accumulation stage, the water eutrophication status is mainly in level IV. Then, with the increase of temperature and rainfall in June and July, algae blooms gradually accumulate so that the water eutrophication status reaches level V.

However, we can see that the deviations appear in eutrophication levels based on measured and simulated values from Table 3 and Figure 5. The reason lies in two aspects: (1) Model approximation: the used WASP model is an approximation of real water quality evolution. This approximation will bring uncertainty and inaccuracy during simulating the evolution process, such as abrupt variation process of eutrophication degree. (2) Model parameter selection: in the simulation, the model parameters are selected as constant values, which can be viewed as a simplification, since the model parameters are time-variant in practice. This simplification will introduce errors when simulating the water quality evolution process.

Here, it should be noted that the modified fuzzy comprehensive evaluation method can be applied to more monitoring stations in Taihu Lake, which will lead to a more comprehensive evaluation result of its eutrophication level. In order to quantitatively represent the accuracy of evaluation results based on simulated values of different methods, the

<table>
<thead>
<tr>
<th>Method</th>
<th>DO (mg/L)</th>
<th>BOD (mg/L)</th>
<th>NH$_3$-N (mg/L)</th>
<th>TN (mg/L)</th>
<th>Chl$_a$ (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial-and-error method</td>
<td>1.47</td>
<td>0.429</td>
<td>0.14</td>
<td>0.52</td>
<td>0.007</td>
</tr>
<tr>
<td>Nonlinear least squares</td>
<td>1.09</td>
<td>0.297</td>
<td>0.12</td>
<td>0.45</td>
<td>0.004</td>
</tr>
<tr>
<td>UKF</td>
<td>0.72</td>
<td>0.215</td>
<td>0.11</td>
<td>0.39</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Figure 2: Results of DO concentration, BOD concentration, NH$_3$-N concentration, TN concentration, and Chl$_a$ concentration.

Table 2: RMSE values of each water quality index obtained by different methods.
Table 3: Standard values of water quality indexes for lakes and reservoirs.

<table>
<thead>
<tr>
<th>Level</th>
<th>SD (m)</th>
<th>BOD (mg/L)</th>
<th>TN (mg/L)</th>
<th>TP (mg/L)</th>
<th>DO (mg/L)</th>
<th>Chl_a (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>10.00</td>
<td>0.18</td>
<td>0.02</td>
<td>0.001</td>
<td>15</td>
<td>0.001</td>
</tr>
<tr>
<td>II</td>
<td>5.00</td>
<td>0.24</td>
<td>0.31</td>
<td>0.004</td>
<td>8.5</td>
<td>0.002</td>
</tr>
<tr>
<td>III</td>
<td>1.50</td>
<td>1.20</td>
<td>1.20</td>
<td>0.023</td>
<td>5.0</td>
<td>0.004</td>
</tr>
<tr>
<td>IV</td>
<td>0.55</td>
<td>6.00</td>
<td>3.60</td>
<td>0.110</td>
<td>2.0</td>
<td>0.01</td>
</tr>
<tr>
<td>V</td>
<td>0.17</td>
<td>15.00</td>
<td>4.70</td>
<td>0.660</td>
<td>1.0</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Figure 3: Triangular linear membership functions of the water quality indexes.

consistency percentage of evaluation results of each method is given in Table 4.

It can be drawn from Table 4 that the accuracy of the evaluation result based on simulated values of UKF is higher than those based on simulated values of trial-and-error method and simulated values of nonlinear least squares.

4.2. Case 2: Beihai Lake

4.2.1. Model Parameter Estimation Result and Analysis. The UKF is also used to estimate the unknown model parameters of Beihai Lake shown in Notations. And the observation matrix is the same as that of Taihu Lake. Then the real-time estimated values of the ten unknown model parameters can be obtained, which are depicted in Figure 6.

The algal blooms formation process of Taihu Lake also applies to Beihai Lake. Following the multistage principle, the average values of unknown model parameters of each stage are shown in Table 5.

Similarly, the estimated values of the unknown parameters in Table 5 are substituted into the water quality soft-sensing method, and the simulation process is carried out. Then Figure 7 depicts the results of DO concentration, BOD concentration, NH₃-N concentration, TN concentration, and Chl_a concentration based on measured values and simulated values of different methods.

From Figure 7, it can be seen that the simulated values are in good agreement with the measured values of each water quality index, which can verify the effectiveness and generality of the fusion water quality soft-sensing method.
Then the model accuracy based on the UKF is better than those based on the nonlinear least squares method and trial-and-error method. Similarly, RMSE is utilized to indicate the deviation between the simulated values and the measured values of each water quality index. By using (25), the specific result is shown in Table 6.

4.2.2. Water Eutrophication Evaluation Result and Analysis. Following the proposed algorithm, the eutrophication evaluation result of Beihai Lake is shown in Figure 8.

<table>
<thead>
<tr>
<th>Method</th>
<th>Trial-and-error method</th>
<th>Nonlinear least squares</th>
<th>UKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency percentage</td>
<td>0.81%</td>
<td>0.84%</td>
<td>0.88%</td>
</tr>
</tbody>
</table>

According to (23), the water eutrophication level based on measured values, simulated values of trial-and-error method, simulated values of nonlinear least squares, and simulated values of UKF is evaluated, and results are shown in Figure 9. It can be seen from Figure 9 that, during the recovery and dormancy stage of algal blooms, the water eutrophication status is basically in level II, while in the biomass increase and accumulation stage, the water eutrophication status is mainly in level IV. Analogously, the modified fuzzy comprehensive evaluation method can be applied to more monitoring stations in Beihai Lake.

In order to quantitatively represent the accuracy of evaluation results based on simulated values of different methods, the consistency percentage is also introduced and utilized for calculation, and the results are shown in Table 7.

It can be drawn from Table 7 that accuracy of evaluation result based on simulated values of UKF is higher than those based on simulated values of trial-and-error method and simulated values of nonlinear least squares, too.
Table 5: Average values of unknown model parameters of each stage.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Symbol</th>
<th>(k_1)</th>
<th>(k_2)</th>
<th>(k_3)</th>
<th>(k_4)</th>
<th>(k_5)</th>
<th>(k_6)</th>
<th>(k_7)</th>
<th>(k_8)</th>
<th>(k_9)</th>
<th>(k_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery</td>
<td></td>
<td>2.681</td>
<td>1.044</td>
<td>0.688</td>
<td>1.285</td>
<td>0.094</td>
<td>−0.047</td>
<td>2.737</td>
<td>1.374</td>
<td>0.001</td>
<td>0.434</td>
</tr>
<tr>
<td>Biomass increase and accumulation</td>
<td></td>
<td>2.695</td>
<td>1.358</td>
<td>2.593</td>
<td>1.066</td>
<td>0.143</td>
<td>0.006</td>
<td>4.652</td>
<td>2.151</td>
<td>0.015</td>
<td>0.673</td>
</tr>
<tr>
<td>Dormancy</td>
<td></td>
<td>2.719</td>
<td>1.568</td>
<td>2.565</td>
<td>0.641</td>
<td>0.167</td>
<td>0.013</td>
<td>4.866</td>
<td>2.548</td>
<td>0.020</td>
<td>0.663</td>
</tr>
</tbody>
</table>

Figure 6: Real-time estimated values of ten unknown model parameters.

5. Conclusions

The fusion water quality soft-sensing method is constructed with a combination between the WASP mechanism model and UKF, and the modified fuzzy comprehensive evaluation method is presented to evaluate the level of water eutrophication. Then, taking Taihu Lake and Beihai Lake as examples, the results show that the simulated values of water quality indexes are in good agreement with the measured values, which can verify the effectiveness and generality of the fusion water quality soft-sensing method, and unknown parameter estimation based on UKF can further improve the accuracy of the model more than nonlinear least squares method and trial-and-error method. Besides, the modified fuzzy comprehensive evaluation method is used to assess the water eutrophication status, and the evaluation results of eutrophication level are consistent in most cases based on simulated values and measured values. Moreover, the modified fuzzy comprehensive evaluation method can be applied to more monitoring stations in Taihu Lake and Beihai Lake, which will lead to a more comprehensive evaluation result of their eutrophication level and provide a scientific reference for water environment management.

In future research, more water quality indexes should be considered in the procedure of water eutrophication evaluation. Furthermore, some quantitative indicators, such as health degree [39, 40], should be introduced to evaluate water eutrophication.
Figure 7: Results of DO concentration, BOD concentration, NH$_3$-N concentration, TN concentration, and Chl$_a$ concentration.

Figure 8: Membership evaluation result of each level based on the measured values and simulated values of UKF.
Table 6: RMSE values of each water quality index obtained by different methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>DO (mg/L)</th>
<th>BOD (mg/L)</th>
<th>NH$_3$-N (mg/L)</th>
<th>TN (mg/L)</th>
<th>Chl$_a$ (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial-and-error method</td>
<td>1.31</td>
<td>0.167</td>
<td>0.07</td>
<td>0.22</td>
<td>0.002</td>
</tr>
<tr>
<td>Nonlinear least squares</td>
<td>0.88</td>
<td>0.151</td>
<td>0.04</td>
<td>0.13</td>
<td>0.001</td>
</tr>
<tr>
<td>UKF</td>
<td>0.68</td>
<td>0.080</td>
<td>0.02</td>
<td>0.09</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 9: Evaluation result of the monitoring station.

Table 7: Consistency percentage of results based on simulated values of different methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Consistency percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial-and-error method</td>
<td>0.91%</td>
</tr>
<tr>
<td>Nonlinear least squares</td>
<td>0.94%</td>
</tr>
<tr>
<td>UKF</td>
<td>0.97%</td>
</tr>
</tbody>
</table>

Notations

$k_1$: Reaeration rate coefficient at 20°C

$k_2$: Deoxygenation rate coefficient at 20°C

$k_3$: Nitrification rate coefficient at 20°C

$k_4$: Sediment oxygen demand coefficient

$k_5$: The maximum growth rate of phytoplankton at 20°C

$k_6$: Phytoplankton mortality (nonzooplankton)

$k_7$: Denitrification rate coefficient at 20°C

$k_8$: Half saturation constant for oxygen limitation of nitrification

$k_9$: Mineralization rate coefficient of dissolved organic nitrogen at 20°C

$k_{10}$: Mineralization rate coefficient of dissolved organic phosphorus at 20°C.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors’ Contributions

Xiaoyi Wang and Zhiyao Zhao conceived and designed the framework of the paper; Jie Jia and Tingli Su performed the experiment; Jiping Xu and Li Wang collected references and processed the data; Xiaoyi Wang and Jie Jia wrote the paper; Zhiyao Zhao revised the paper.

Acknowledgments

This work was financially supported by the National Key R&D Program of China (2017YFC1600605), the National Natural Science Foundation of China (61703008), and the Research Foundation for Youth Scholars of Beijing Technology and Business University (QNJJ2018-18).

References


