Water Quality Assessment Based on Probabilistic Echo State Networks

Li Dongping, Yang Yingchun and Wen Jin*

***Abstract*—Water quality evaluation is important for water environment management and protection. The conventional neural network based evaluation algorithm only assesses water quality level , but it can not offer us to what extend the water quality belongs to the corresponding level. To solve this problem, the water quality assessment by probabilistic echo state networks (PESN) is proposed. PESN provides the pollution level in the form of probability output, which offers quantificational similarity to each water quality level. PESN is consisted of sigmoid posterior probability mapping and pairwise coupling. Firstly the *M* classification is divided into *M*(*M*-1)/2 pairs of two classification. Secondly each two classification ESN is transformed to probabilistic output by sigmoid mapping. Thirdly probabilistic output is calculated through pairwise coupling, which is used to fuse all the two classification probabilistic outputs. The water quality estimation case study shows that, PESN provides evaluation result in the probability form. PESN probability can indicate the pollution degree, while ESN lacks such ability. Compared with probabilistic support vector machine (PSVM), the number of user specified parameters for PESN is less than PSVM, which reduces PESN computing time.**

***Index Terms*—water pollution, water quality prediction, neural network, sigmoid function, pairwise coupling tracking**

I. Introduction

Water quality assessment is a prerequisite to prevent and control water pollution. Meanwhile water quality estimation is the indispensable technology to rationally develop and make full use of water resources. There exists many water quality assessment methods, such as single factor evaluation, fuzzy comprehensive evaluation, neural network evaluation, etc.

Neural networks are popular tools for water quality evaluation with its high nonlinear mapping ability, self-learning ability and parallel information processing ability [1,2]. Literature [3] designs a feed-forward neural network model for estimating the water quality index of Kinta River (Malaysia). The experiment result shows its good performance in water quality index calculation and forecasting. To estimate uncertainty in any water quality risk evaluation, literature [4] presents a monte carol simulation with artificial neural network. Simulation test indicates the applicable of the proposed method to assess water quality risks. In addition, wavelet neural network[5], genetic algorithms based neural network[6], feed forward error back propagation based neural network[7] are applied to water quality assessment separately.

However, the afore mentioned algorithms only provide water quality classification label, namely the predicted sample belongs to which water quality level. They lack the ability to give probability, i.e., to what extend the predicted sample is similar to different water quality level. The form of probability output is more suitable than only offering water quality level label, because the probability form has the ability to provide us how serious the pollution is, especially when two samples have the same water quality level.

To solve this problem, a novel neural network learning algorithm, echo state network (ESN) [8-19], is improved to compute probability for water quality evaluation. The proposed probabilistic ESN (PESN) is developed with the combination of sigmoid posterior probability function mapping and pairwise coupling. The presented PESN is applied to evaluate the water quality of Dianchi lake in Yunnan province, China. The test result indicates the effectiveness of PESN.

The rest sections are organized as follows. In section 2, the standard ESN is briefly reviewed. Section 3 details the proposed PESN, while section 4 describes the water quality evaluating algorithm based on PESN. The case study of Dianchi lake is performed in section 5. Finally the conclusion is given in section 6.

II. Brief review of ESN

ESN state equations are as follows,

$$\boldsymbol{u}(k+1) = f\left(W_{in}\boldsymbol{x}(k+1) + W_u\boldsymbol{u}(k)\right) \tag{1}$$

$$\boldsymbol{t}(k+1) = \boldsymbol{\beta}\boldsymbol{u}(k) \tag{2}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^M$, $\boldsymbol{u}(k) \in \mathbb{R}^L$ and $\boldsymbol{t}(k) \in \mathbb{R}$ represent input variable, state variable and output variable separately. k is the current time instant. $W_{in} \in \mathbb{R}^{L\times M}$, $W_u \in \mathbb{R}^{L\times L}$ and $\boldsymbol{\beta} \in \mathbb{R}^L$ represent input weight matrix, internal link matrix and output weight matrix respectively. W_{in} and W_u are generated randomly and keep constant, while $\boldsymbol{\beta}$ is computed via training data.

ESN regression model with L nodes in the internal layer can be written in matrix form as,

$$\boldsymbol{H\beta} = \boldsymbol{T} \tag{3}$$

where

$$\boldsymbol{H} = [\boldsymbol{u}(1)\ \boldsymbol{u}(2)\ \cdots\ \boldsymbol{u}(N)]^{\mathrm{T}} \tag{4}$$

Manuscript received November 27, 2016; revised February 18, 2017.

Li Dongping is with institute of Information Science and Technology, Kunming University, 650214, China (e-mail: 164011632@qq.com).

Yang Yingchun is with China Telecom Corporation Limited Yunnan Branch, Kunming, 650200, China (gddx89@21cn.com).

Wen Jin is the corresponding author (*), who is in the institute of Information Science and Technology, Kunming University, 650214, China (e-mail: 1003361829@qq.com).

$$\boldsymbol{T}=[\boldsymbol{t}(1)\ \boldsymbol{t}(2)\ \cdots\ \boldsymbol{t}(N)]^{\mathrm{T}} \tag{5}$$

where N is the number of training samples. The output weights $\boldsymbol{\beta}$ can be calculated by

$$\hat{\boldsymbol{\beta}}=\boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{6}$$

where $\boldsymbol{H}^{\dagger}$ is a Moore-Penrose pseudo-inverse of $\boldsymbol{H}$[20]. Supposing $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ is nonsingular matrix, then $\boldsymbol{H}^{\dagger}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}$. Thus $\hat{\boldsymbol{\beta}}$ in equation (6) can be computed as [21]

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T} \tag{7}$$

Seen from equation (7), it needs to assume $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ nonsingular when calculating $\boldsymbol{\beta}$. To avoid singular matrix, equation (3) can be replaced via seeking $\boldsymbol{\beta}$ through optimizing[22]

$$\min_{\boldsymbol{\beta}}\{\|\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{T}\|^2+\lambda\|\boldsymbol{\beta}\|^2\} \tag{8}$$

where $\|\cdot\|$ is 2-norm, $\lambda>0$ is a regularization factor.

The optimization value of equation (8) is given by [22]

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+\lambda\mathbf{I})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T} \tag{9}$$

For two classification problem, decision equation is as follow,

$$f(\boldsymbol{x})=\mathrm{sign}\left(h(\boldsymbol{x})\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+\lambda\mathbf{I}\right)^{-1}\boldsymbol{T}\right) \tag{10}$$

For multi classification problem, decision equation is as follow,

$$label(\boldsymbol{x})=\arg\max_{i\in\{1,\cdots,m\}} f_i(\boldsymbol{x}) \tag{11}$$

where $f_i(\boldsymbol{x})$ is the ith output node, and $f(\boldsymbol{x})=\left[f_1(\boldsymbol{x}),\cdots,f_m(\boldsymbol{x})\right]^{\mathrm{T}}$.

III. PROBABILISTIC ECHO STATE NETWORK

A. Probabilistic Mapping by Sigmoid Function

For classification problem, ESN only provides output classification label, namely to which classification the predicted sample belongs to. ESN lacks the ability to give probability, i.e., the probabilities of the predicted sample are similar to different classifications. In order to transform the ESN output into the probability form, posterior probability mapping method based on sigmoid function is proposed by Platt[23]. The mapping function of two classification problem {+1, -1} is expressed as follows,

$$P(t=+1\,|\,\boldsymbol{x})=\frac{1}{1+\exp(Af_{+1}(\boldsymbol{x})+B)} \tag{12}$$

where +1 and -1 are two type classification label; $f_{+1}(\boldsymbol{x})$ represents the output corresponding to classification label +1; $P(t=-1\,|\,\boldsymbol{x})=1-P(t=+1\,|\,\boldsymbol{x})$. To determine A and B, the log likelihood function needs to be minimized via training samples $\{(\boldsymbol{x}_l,\boldsymbol{t}_l)\}_{l=1}^{N}$.

$$\min_{A,B}:\ -\sum_{l=1}^{N}t_l\log(p_l)+(1-t_l)\log(1-p_l) \tag{13}$$

where

$$p_l=\frac{1}{1+\exp(Af_{+1}(\boldsymbol{x}_l)+B)} \tag{14}$$

$$t_l=\begin{cases}\dfrac{N_+ +1}{N_+ +2} & \text{if } t_l=+1\\[2mm] \dfrac{1}{N_- +2} & \text{if } t_l=-1\end{cases} \tag{15}$$

where N_+ and N_- are training samples number of label +1 and -1 separately. Levenberg Marquardt algorithm[23] is applied to calculate A and B.

To prove the rationality of probabilistic output by sigmoid mapping, the method according to reference [23] is used. This method is to compare the fitting degree of posterior probability between sigmoid mapping and bayesian criterion. Taking adult data in UCI database for example, the probability $P(t=+1|\boldsymbol{x})$ of sigmoid mapping and bayesian criterion is shown in figure 1, it can be seen that the fitting degree is high. As a result, it certifies the effectiveness of sigmoid mapping method.

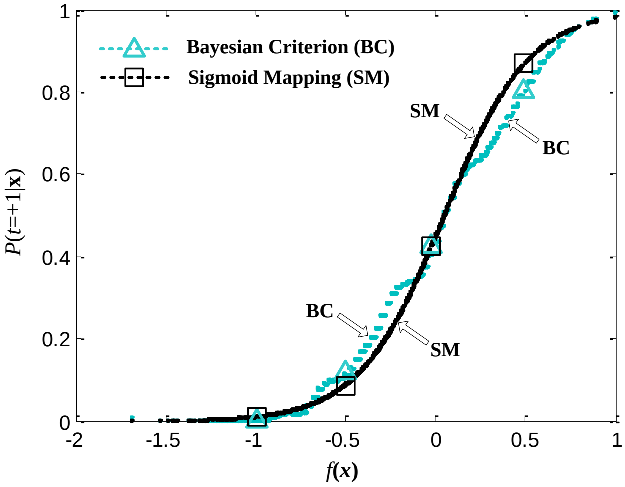


Fig. 1 The probability $P(t=+1|\boldsymbol{x})$ of sigmoid mapping and bayesian criterion

B. Pairwise Coupling

Posterior probability mapping method based on sigmoid function is only suitable for two classification problem. To solve M types classification problem, pairwise coupling algorithm is utilized to dived the M types classification into $M(M-1)/2$ two types classification, where $M>2$. For the two classification situation composed of the ith and jth classification, r_{ij} represents the probability of input sample $\boldsymbol{x}$ belongs to the ith classification.

$$r_{ij}=P_{ij}^{\mathrm{ESN}}(t=i\,|\,t=i\text{ or } j,\boldsymbol{x}) \tag{16}$$

where $P_{ij}^{\mathrm{ESN}}(\cdot)$ is ESN probability output function via equation (12). $P_{ij}^{\mathrm{ESN}}(\cdot)$ is obtained through training samples consisted of the ith and jth classification, $i,j=1,2,\cdots,M$, $i\neq j$.

To compute input vector $\boldsymbol{x}$ belongs to the ith classification probability p_i, $M(M-1)/2$ two types classification probabilities r_{ij} need to be fused, where $p_i=P(t=i\,|\,\boldsymbol{x})$, $i=1,2,\cdots,M$. Therefore to obtain p_i is equivalent to minimize the following equation [24].

$$\begin{gathered}\min_{\boldsymbol{p}=\{p_1,\cdots,p_D\}}:\sum_{i=1}^{M}\sum_{j=1;j\neq i}^{M}\left(r_{ji}p_i-r_{ij}p_j\right)^2\\ \text{subject to }\sum_{i=1}^{M}p_i=1,\ p_i\geq 0,\ \forall i.\end{gathered} \tag{17}$$

Equation (17) can be rewritten as the following objective function,

$$\min_{\boldsymbol{p}=[p_1,\cdots,p_M]^{\mathrm{T}}} 2\boldsymbol{p}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{p} \equiv \min_{\boldsymbol{p}=[p_1,\cdots,p_M]^{\mathrm{T}}} \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{p} \tag{18}$$

where

$$Q_{ij} = \begin{cases} \sum_{s=1;s\neq i}^{M} r_{si}^2 & \text{if } i=j \\ -r_{ji}r_{ij} & \text{if } i\neq j \end{cases} \tag{19}$$

The objective function (18) can be divided into a positive factor of four, as a result, the objective function is a simple form $\boldsymbol{Qp}$. Seen from equation (19), $\boldsymbol{Q}$ is positive semi-definite since for any $\boldsymbol{v}\neq\boldsymbol{0}$,

$$\boldsymbol{vQv} = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}(r_{ji}v_i - r_{ij}v_j)^2 \geq 0 \tag{20}$$

Thus except the constraint $p_i \geq 0$, equation (18) can be treated as a linear equality constrained convex quadratic programming problem. Then by adding a scalar $\boldsymbol{b}$,

$$\begin{bmatrix} \boldsymbol{Q} & \boldsymbol{e} \\ \boldsymbol{e}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{1} \end{bmatrix} \tag{21}$$

where $\boldsymbol{Qp}$ is the derivative of equation (18), $\boldsymbol{b}$ is the Lagrange multiplier of the equality constraint $\sum_{i=1}^{M} p_i = 1$, $\boldsymbol{e}$ represents unit vector with the dimension of M, $\boldsymbol{0}$ is zero vector with the dimension of M. Then minimizing function (18) is translated into computing equation (21). Here Cholesky factorization is applied to compute equation (21). At last, the probability $\boldsymbol{p}$ can be obtained by the followings.

$$\boldsymbol{p} = -\boldsymbol{b}\boldsymbol{Q}^{-1}\boldsymbol{e} \tag{22}$$

where $\boldsymbol{p} = \{p_1,\cdots,p_i,\cdots,p_M\}$, where M is the number of classification labels, $\boldsymbol{b}$ is computed by

$$\boldsymbol{b} = -\frac{1}{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{e}} \tag{23}$$

C. PESN Algorithm Flow

Given training samples $\{(\boldsymbol{x}_l,\boldsymbol{t}_l)\}_{l=1}^{N}$, where $\boldsymbol{x}_l \in \mathbb{R}^n$ is input vectors and $t_l \in \{1,\cdots,M\}$ is classification label. Then PESN algorithm flow is organized as follows.

Step 1 $\{(\boldsymbol{x}_l,\boldsymbol{t}_l)\}_{l=1}^{N}$ are divided into $M(M$-1)/2 groups two classification training samples $\{i, j\}$, where training samples $\{(\boldsymbol{x}_l,\boldsymbol{t}_l^{'})\}_{l=1}^{N_{ij}}$ are consisted of label i and j, $t_l^{'} \in \{i, j\}$. If $t_l^{'}=i$, let $\boldsymbol{t}_l^{*}=1$; if $t_l^{'}=j$, let $\boldsymbol{t}_l^{*}=-1$. Then $\{i, j\}$ training samples are changed as $\{(\boldsymbol{x}_l,\boldsymbol{t}_l^{*})\}_{l=1}^{N_{ij}}$, $i, j=1,2,\cdots,M$, $i\neq j$.

Step 2 $\{(\boldsymbol{x}_l,\boldsymbol{t}_l^{*})\}_{l=1}^{N_{ij}}$ are substituted into equation (9) to compute the corresponding ESN output weight $\boldsymbol{\beta}^{ij}$. Then ESN $f^{ij}(\boldsymbol{x}_l)$ of $\boldsymbol{x}_l$ can be computed. $\{f^{ij}(\boldsymbol{x}_l)\}_{l=1}^{N_{ij}}$ are used to calculate $\{i, j\}$ posterior probability mapping parameters A^{ij} and B^{ij} by equation (14) and (15)， $l=1,2,\cdots,N_{ij}$ 。

Step 3 For under estimated input vector $\boldsymbol{x}$, ESN output $f^{ij}(\boldsymbol{x})$ and probability output r_{ij} corresponding to $\{i, j\}$ classification are computed, where $i, j=1,2,\cdots,M$, $i\neq j$.

$$f^{ij}(\boldsymbol{x}) = h(\boldsymbol{x})\boldsymbol{\beta}^{ij} \tag{24}$$

$$r_{ij} = \frac{1}{1+\exp(A^{ij}f^{ij}(\boldsymbol{x})+B^{ij})} \tag{25}$$

Step 4 r_{ij} is substituted into equation (19) to compute $\boldsymbol{Q}$. Then $\boldsymbol{Q}$ is substituted into equation (22) to compute the probability $\{p_1,\cdots,p_M\}$.

IV. WATER QUALITY ASSESSMENT BY PESN

Based on the proposed PESN, water quality predicting algorithm is presented as follows.

Step 1 The water quality parameters which can indicate pollution are selected.

Step 2 The known different water quality level data are arranged to form training samples. PESN model is built through the training samples.

Step 3 The water quality parameters under assessed are taken into the trained PESN. Then PESN provides the probabilities the water belongs to different pollution level.

V. CASE STUDY

In this section, the performance of the proposed PESN algorithm are tested using water quality data from Dianchi lake. To verify PESN effectiveness, the assessment result and time consuming of PESN is compared with standard ESN and probabilistic SVM (PSVM). All the simulations are carried out in MATLAB R2011b environment running in an Intel Core i3, 3.30 GHz CPU and a 2 GB RAM.

A. Study Area and Data

Dianchi lake is the largest freshwater lake in Yunnan province, China. Dianchi lake has an area of approximately 330 km^2 and its altitude is 1886m. Because Dianchi lake is the only water area to take up polluted water in Kunming city, Yunnan province, the lake has been seriously polluted since 1990s. And the main pollution type is eutrophication due to the discharge of industrial waste water.

Fig. 2 Water quality monitoring stations on Dianchi lake

The experiment data is collected from eight water quality monitoring stations on Dianchi lake (see figure 2). The original water quality dataset is obtained on six months every year (dry season: January, March, November, rainy season: May, July, September) during 2010 to 2014.

Total phosphorus (TP), oxygen consumption (OS), water transparency (WT) and total nitrogen (TN) are chosen as the water quality parameters since they are main pollution indicators for Dianchi lake. The corresponding pollution classification standard is shown in table I.

In order to build water quality classification model by PESN, training samples should be generated according to pollution classification standard, as shown in table I. For each pollution level, four water quality parameters (TP, OS, WT, TN) are randomly generated from the defined range. Taking "light pollution" level for example, TP, OS, WT, TN are randomly selected from the range [1 4], [0.09 0.36], [12 37], [0.02 0.06] separately. By using the mentioned method, 100 training samples for each pollution level are generated. As a result, the total number of training samples is 500. To evaluate the performance of the proposed algorithm, 240 water quality data are used as testing samples, consist of eight water quality monitoring stations, each station with six data every year from 2010 to 2014.

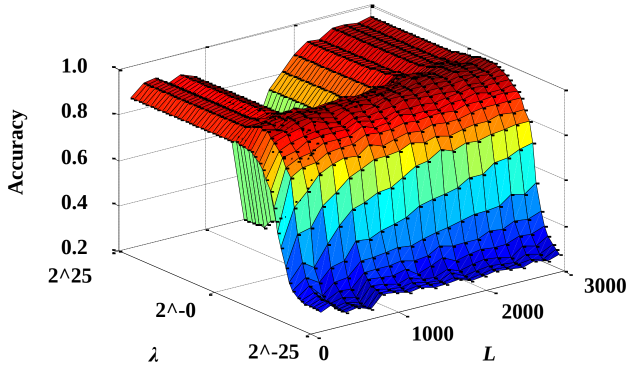
B. *User Specified Model Parameters*

It is known that the prediction accuracy of PSVM depends mainly on model parameters (C, γ), where C is cost parameter and γ is kernel parameter separately. For ESN and PESN, parameters (L, λ) are user specified, where L is the internal neuron number and λ is regularization factor. The high prediction accuracy of PSVM, ESN and PESN are commonly obtained between a very narrow range of such model parameters combination. Therefore the optimal model parameters combination of (C, γ) for PSVM, and (L, λ) for ESN and PESN should be selected. The 500 training data are separated into two sub-datasets: 400 randomly chosen training data for initial training, and the other 100 data for parameters optimization (named validating data).

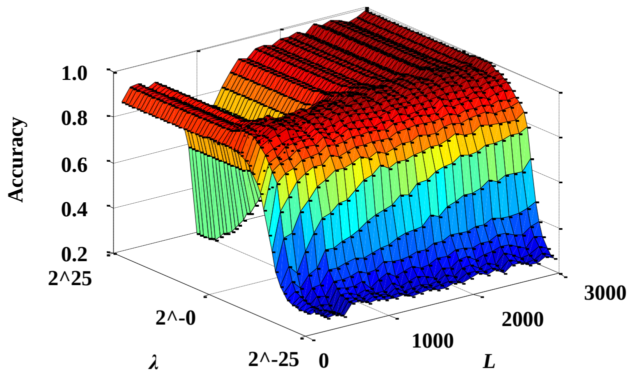
1. *Model parameters affection for ESN, PESN and PSVM*

For ESN and PESN, the user specified parameter L is selected from the range $\{100, 200, \ldots, 3000\}$, while λ is selected from the range $\{2^{-24}, 2^{-8}, \ldots, 2^{9}, 2^{25}\}$. PESN validating data prediction accuracy to different (L, λ) is shown in figure 3(a). It can be seen from figure 3(a) that PESN prediction accuracy is high as long as internal neuron number L is large enough, namely the prediction accuracy is not sensitive to L. As a result, L can be set as a constant, and the only parameter need to be specified is regularization factor λ for PESN. Validating data prediction accuracy to different (L, λ) based on ESN is shown in figure 3(b). Similar to PESN, ESN prediction performance depends mainly on λ.

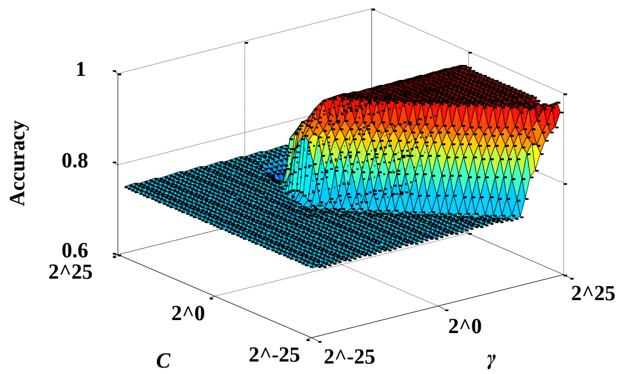
For PSVM, the user specified parameters (C, γ) are selected from the range $\{2^{-24}, 2^{-8}, \ldots, 2^{9}, 2^{25}\}$. The validating data prediction accuracy to different (C, γ) is shown in figure 3(c). Seen from figure 3(c), both C and γ affect PSVM prediction accuracy. Unlike internal neuron number L of ESN and PESN, which can be set to be constant, both C and γ should be optimized according to validating data prediction accuracy.



(a) PESN



(b) ESN



(c) PSVM

Fig. 3 Validating data accuracy of different methods for different parameters

2. *Model parameters optimizing result*

For ESN and PESN parameters optimizing, L is fixed as 2500, λ is chosen from the range $\lambda = \{2^{-24}, 2^{-8}, \ldots, 2^{9}, 2^{25}\}$. For PSVM, ($C$, γ) are chosen from the range $C = \{2^{-24}, 2^{-8}, \ldots, 2^{9}, 2^{25}\}$, $\gamma = \{2^{-24}, 2^{-8}, \ldots, 2^{9}, 2^{25}\}$. The optimal parameters are chosen as the ones with the highest validation accuracy.

TABLE I
POLLUTION CLASSIFICATION STANDARD

Pollution level	Total phosphorus (TP) /(μg° L^{-1})	Oxygen consumption (OS) /(mg° L^{-1})	Water transparency (WT) /m	Total nitrogen (TN) /(mg° L^{-1})
Clean (I)	0-1	0-0.09	70-37	0-0.02
Light pollution (II)	1-4	0.09-0.36	37-12	0.02-0.06
Moderate pollution (III)	4-23	0.36-1.80	12-2.4	0.06-0.31
Heavy pollution (IV)	23-110	1.80-7.10	2.4-0.55	0.31-1.20
Serious pollution (V)	110-660	7.10-27.10	0.55-0.17	1.20-4.60

Finally the user specified parameters for different algorithms are shown in Table II.

TABLE II
MODEL PARAMETERS OPTIMIZATION

PSVM	ESN	PESN
$C = 2^{-2}, \gamma = 2^{22}$	$L = 2500, \lambda = 2^{5}$	$L = 2500, \lambda = 2^{1}$

C. *Water Quality Assessment Result*

The testing data collected from water quality monitoring stations on Dianchi are applied to estimate water quality. The measured water quality values (TP, OS, WT, TN) on March and September, 2012 are listed in table III, while the water quality level prediction by ESN, PSVM and PESN is shown in table IV.

Seen from table III and table IV, Water quality level of eight monitoring stations are between IV (heavy pollution) and V (serious pollution) on March and September, 2012. The result indicates high pollution level for Dianchi lake. ESN simply provides the water quality level assessment without to what extend the water quality belongs to the corresponding pollution level. As contrast, PSVM and PESN give the assessment result in the form of probability, i.e., the probability represents the membership degree for each pollution level. Taking testing data at monitoring station A and B on March, 2012 for example, water quality assessment level is V by ESN, PSVM and PESN. Obviously the pollution degree at station B is higher than station A. But ESN lacks the ability to show such information. ESN indicates the pollution degree at station A and B is the same, which misleads the water quality estimation result. PESN provides the pollution level probability as 0.6208 for level V and 0.3511 for level IV at station A, meanwhile 0.9250 for level V and 0.0649 for level IV at station B. From the PESN estimation, the conclusion that pollution degree at station B is more serious than station A can be obtained. PSVM could provides the similar result as PESN.

TABLE III
THE MEASURED WATER QUALITY VALUES

Date	Station	Total phosphorus (TP) /(μg° L^{-1})	Oxygen consumption (OS) /(mg° L^{-1})	Water transparency (WT) /m	Total nitrogen (TN) /(mg° L^{-1})
Mar, 2012	A	151	6.23	0.72	2.1
	B	282	12.14	0.41	5.5
	C	911	6.31	0.81	16.1
	D	480	9.53	0.53	7.9
	E	123	6.22	0.37	1.8
	F	91	5.02	0.63	1.7
	G	201	9.61	0.39	2.2
	H	252	4.23	0.58	1.9
Sep, 2012	A	165	6.50	0.79	3.9
	B	322	29.50	0.26	9.8
	C	281	13.85	0.18	12.2
	D	330	14.51	0.19	56
	E	112	7.13	0.59	1.5
	F	85	7.32	0.60	2.1
	G	152	7.11	0.56	2.3
	H	241	8.01	0.51	3.8

TABLE IV
WATER QUALITY LEVEL PREDICTION

Date	Station	ESN	PSVM					PESN				
			I	II	III	IV	V	I	II	III	IV	V
Mar, 2012	A	V	0.0001	0.0037	0.0238	0.3412	0.6312	0.0000	0.0095	0.0186	0.3511	0.6208
	B	V	0.0003	0.0043	0.0075	0.0814	0.9065	0.0002	0.0039	0.0060	0.0649	0.9250
	C	V	0.0000	0.0001	0.0005	0.0022	0.9972	0.0000	0.0001	0.0004	0.0010	0.9985
	D	V	0.0001	0.0004	0.0009	0.0074	0.9912	0.0000	0.0002	0.0007	0.0043	0.9948
	E	V	0.0002	0.0115	0.0149	0.3318	0.6416	0.0000	0.0067	0.0152	0.3413	0.6368
	F	IV	0.0002	0.0088	0.0169	0.5669	0.4072	0.0001	0.0073	0.0166	0.5671	0.4088
	G	V	0.0002	0.0085	0.0166	0.2665	0.7083	0.0001	0.0082	0.0156	0.2393	0.7368
	H	V	0.0001	0.0052	0.0089	0.1009	0.8850	0.0005	0.0056	0.0105	0.1250	0.8585
Sep, 2012	A	V	0.0001	0.0140	0.0312	0.3433	0.6114	0.0001	0.0140	0.0312	0.3371	0.6176
	B	V	0.0002	0.0007	0.0011	0.0113	0.9866	0.0001	0.0012	0.0016	0.0164	0.9807
	C	V	0.0003	0.0030	0.0049	0.0507	0.9411	0.0001	0.0025	0.0037	0.0383	0.9554
	D	V	0.0003	0.0015	0.0021	0.0225	0.9736	0.0002	0.0020	0.0029	0.0301	0.9647
	E	V	0.0001	0.0071	0.0161	0.3658	0.6109	0.0002	0.0073	0.0069	0.3848	0.6008
	F	IV	0.0003	0.0092	0.0168	0.5487	0.4250	0.0002	0.0077	0.0173	0.5590	0.4158
	G	V	0.0002	0.0137	0.0313	0.3716	0.5831	0.0001	0.0139	0.0313	0.3627	0.5921
	H	V	0.0002	0.0076	0.0146	0.2145	0.7631	0.0001	0.0073	0.0137	0.1923	0.7867

The other 224 measured water quality samples are applied to ESN, PSVM and PESN separately. The predicting result is similar to table IV. The total time consuming for assessment 240 testing samples is shown in table V. Table V indicates that ESN time consuming is the lowest. For PESN, additional pairwise coupling process should be performed, so time consuming of PESN is larger than ESN. PESN time consuming is lower than PSVM, since the number of user specified parameter for PESN is only one while PSVM is two.

TABLE V
TIME CONSUMING OF THREE ALGORITHMS (S)

ESN	PSVM	PESN
1.1252	3.1547	2.1616

VI. CONCLUSIONS

This study describes the application of PESN to a water quality evaluation problem. ESN lacks the ability to provide assessment in probability form, which form can quantitatively indicates the water quality. PESN based method overcomes such shortcoming. Experiment shows that PESN can provide water pollution level in the form of probability, which is more clear than ESN to show different pollution degree when pollution level is same. Though PSVM also has the ability to provide probability, its user specified parameter is one more than PESN. Therefore, PESN is recommended for water quality evaluation.

REFERENCES

[1] Li H-y, Teng J, Li Z-h, Zhang L, "Nonlinear Dynamic Analysis Efficiency by Using a GPU Parallelization," *Engineering Letters*, vol. 23, no. 4, pp 232-238, 2015.

[2] Reis TSd, Gomide W, Anderson JADW, "Construction of the Transreal Numbers and Algebraic Transfields," *IAENG International Journal of Applied Mathematics*, vol. 46, no. 1, pp 11-23, 2016.

[3] Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, Ramli MF, "Artificial neural network modeling of the water quality index for Kinta River," *Marine Pollution Bulletin*, vol. 64, no. 1, pp 2409-2420, 2012.

[4] Jiang Y, Nan Z, Yang S, "Risk assessment of water quality using Monte Carlo simulation and artificial neural network method," *Journal of Environmental Management*, vol. 122, no. 1, pp 130-136, 2013.

[5] Xu L, Liu S, "Study of short-term water quality prediction model based on wavelet neural network," *Mathematical & Computer Modelling*, vol. 58, no. s 3–4, pp 807-13, 2013.

[6] Ding YR, Cai YJ, Sun PD, Chen B, "The Use of Combined Neural Networks and Genetic Algorithms for Prediction of River Water Quality," *Journal of Applied Research & Technology*, vol. 12, no. 3, pp 493-499, 2014.

[7] Sarkar A, Pandey P, "River Water Quality Modelling Using Artificial Neural Network Technique," *Aquatic Procedia*, vol. 4, no.1, pp. 1070-1077, 2015.

[8] Bianchi FM, Scardapane S, Uncini A, Rizzi A, Sadeghian A, "Prediction of telephone calls load using Echo State Network with exogenous variables," *Neural Networks*, vol. 71, no. 1, pp 204-213, 2015.

[9] Morando S, Jemei S, Hissel D, Gouriveau R, Zherouni N, "Predicting the Remaining Useful Lifetime of a Proton Exchange Membrane Fuel Cell using an Echo State Network," *International Discussion on Hydrogen Energy & Applications*, vol.1, no.1, pp 1-9, 2014.

[10] Feng P, Hai-bo Z, "Online Sequential Extreme Learning Machine Based Multilayer Perception with Output Self Feedback for Time Series Prediction," *Journal of Shanghai Jiaotong University*, vol. 18, no. 3, pp 366-375, 2013.

[11] Lun SX, Lin J, Yao XS, "Time series prediction with an improved echo state network using small world network," *Acta Automatica Sinica*, vol. 41, no. 9, pp 1669-1679, 2015.

[12] Lun SX, Yao XS, Qi HY, Hu HF, "A novel model of leaky integrator echo state network for time-series prediction," *Neurocomputing*, vol. 159, no. 3, pp 58-66, 2015.

[13] Xu X, Niu D, Fu M, Xia H, Wu H, "A Multi Time Scale Wind Power Forecasting Model of a Chaotic Echo State Network Based on a Hybrid Algorithm of Particle Swarm Optimization and Tabu Search," *Energies*, vol. 8, no. 11, pp 12388-12408, 2015.

[14] Huang B, Qin G, Zhao R, Wu Q, Shahriari A, "Recursive Bayesian echo state network with an adaptive inflation factor for temperature prediction," *Neural Computing and Applications*, pp 1-9, 2016.

[15] Huang J, Qian J, Liu L, Wang Y, Xiong C, Ri S, "Echo state network based predictive control with particle swarm optimization for pneumatic muscle actuator," *Journal of the Franklin Institute*, vol. 353, no. 12, pp 2761-82, 2016.

[16] Morando S, Jemei S, Hissel D, Gouriveau R, Zerhouni N, "Proton exchange membrane fuel cell ageing forecasting algorithm based on Echo State Network," *International Journal of Hydrogen Energy*, 2016.

[17] Wang L, Wang Z, Liu S, "An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm," *Expert Systems with Applications*, vol. 43, no. C, pp 237-249, 2016.

[18] Yao W, Zeng Z, Lian C, "Generating probabilistic predictions using mean-variance estimation and echo state network," *Neurocomputing*, pp 1-7, 2016.

[19] Zhao Y, Gao H, Beaulieu N, Chen Z, Ji H, "Echo State Network for Fast Channel Prediction in Ricean Fading Scenarios," *IEEE Communications Letters*, vol. PP, no. 99, pp 1-2, 2016.

[20] Guang-bin H, Qin-yu Z, Chee Kheong S, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp 489-501, 2006.

[21] Hai W, Gang Q, Xiang-qian F, "Predicting consumer sentiments using online sequential extreme learning machine and intuitionistic fuzzy sets," *Neural Computing & Applications*, vol. 22, no. 3, pp 479-489, 2013.

[22] Hieu Trung H, Yonggwan W, "Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 1930-1935, 2011.

[23] John C P, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers: Massachusetts Institute of Technology Press*, pp 1-11, 1999.

[24] Suprapto, Wardoyo R, Widjaja BH, Pulungan R, "A Formal Proof of Correctness of Construct Association from PROMELA to Java," *IAENG International Journal of Computer Science*, vol. 42, no. 4, pp 313-331, 2015.

Li Dongping, female, was born in the city of Kunming, Yunnan Province, China on 1971. She acts as an associate professor in the institute of Information Science and Technology, Kunming University. Her interest is data mining.

Yang Yingchun (1971.02-), male, acts as a senior engineer in China Telecom Corporation Limited Yunnan Branch, Kunming. His interest is data analysis.

Wen Jin (1963.04-), corresponding author, male, acts as an associate professor in the institute of Information Science and Technology, Kunming University.