Diagnosing reservoir water quality using self-organizing maps and fuzzy theory

Ruei-Shan Lu*, Shang-Lien Lo

Graduate Institute of Environmental Engineering, National Taiwan University, 71 Chou-Shan Rd., Taipei 106, Taiwan, ROC

Abstract

Since trophic status assessment of water quality is very important for the water resources management, the assessment results obtained from using only one parameter may easily mislead or bias the decision makers or managers. Even when using a multivariable index, how to determine the weights of all factors is debatable. In this research, one complementary evaluation method, self-organizing map (SOM), for diagnosing water quality has been used to develop a trophic state classifier and is illustrated with a case study of trophic status assessment for Fei-Tsui Reservoir in Taiwan. The historical database was collected from the management agency of Fei-Tsui Reservoir from 1987 to 1995. The results of SOM are compared with those of the Carlson index and Fuzzy synthetic evaluation, showing that the inconsistent records can be mapped to the conflicting data zone of the SOM output map. In addition, SOM creates a diagnostic axis on the map to express the trophic status of the water body. As long as the SOM model is well-trained, new records can be assessed and classified as either one of three trophic levels or special cases. If special water quality conditions are expressed on the SOM output, those data can reveal that either total phosphorus (TP) or chlorophyll A (chl *a*) is higher than usual. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Self-organizing map (SOM); Eutrophication; Reservoir water quality; Fuzzy theory

1. Introduction

Lush vegetation, a high concentration of salts, and high turbidity are the defining characteristics of eutrophication. In a single-variable index, a physico-chemical parameter that is representive of these characteristics is chosen, and its threshold value is set as the trophic standard. Application of this method often produces inconsistent results when different parameters are chosen. Since trophic status assessment of water quality is very important for the water resources management, the assessment results obtained from using only one parameter may easily mislead or bias the user. For this reason, the multivariable trophic state indexing

methods were developed. The most commonly used multivariable indices are the Carlson [1], Morihiri [2], and North Carolina indices [3].

Multivariable indexing allows for a more thorough investigation of water quality and a more continuous description of the eutrophication process, but several problems are inherent in this method. The first problem is that geographical and atmospheric factors influence eutrophication, causing the standards for each area to differ. Consequently, the three aforementioned indices are heavily regional in nature. Another problem involves unreasonable classification standards. For example, the Carlson index gives TSI = 49 and 50 different classifications but TSI = 41 and 49 the same. A further problem with multivariables is that it gives different weights to each parameter. Thus, the scientific community has been unable to agree upon a single, reliable trophic state index.

*Corresponding author. Tel.: +886-2-236-25373; fax: +886-2-239-2881.

E-mail address: rslu@ms.cc.ntu.edu.tw (R.-S. Lu).

Classification of river water quality using fuzzy theory began in the 1990s. Kung et al. [4] published the application of fuzzy clustering analysis to water quality classification. Lu et al. [5] used fuzzy synthetic evaluation to analyze and compare the water quality of Fei-Tsui Reservoir in Taiwan. The results of this investigation showed that most of the water quality records were evaluated as the same trophic state by both the Carlson index and Fuzzy synthetic evaluation. However, the trophic states of several data were assessed inconsistently. These data were termed "conflicting data". During the first 5 years after the reservoir began operation, the water quality of the Fei-Tsui Reservoir was unstable, which was caused by a short-term overturn effect from 1987 to 1992. It was found that the short-term change of water quality and the overturn phenomena could not be observed with the Carlson index from 1987 to 1992 but could be expressed by fuzzy synthetic evaluation.

The self-organizing map (SOM) is a neural-network model and algorithm that implements a characteristic nonlinear projection from the high-dimensional space of sensory or other input signals onto a low-dimensional array of neurons [6]. The term "self-organizing" refers to the ablity to learn and organize information without being given the associated-dependent output values for the input pattern [7]. The SOM is able to map a structured, high-dimensional signal manifold onto a much lower dimensional network in an orderly fashion. The network organizes itself by adjusting the synaptic weights as the input patterns are presented to it; hence, discovery of a new pattern is possible at any instant. Moreover, SOM is noise tolerant; this property is highly desirable when site-measured data are used.

Interesting SOM applications have been reported in mainly three fields: exploratory data analysis or data mining, the identification and monitoring of complex process states, and pattern classification [8].

This study applies SOM to the classification of reservoir water quality for the same case study reported previously [5]. The results of cross-validation between the Carlson index, fuzzy synthetic evaluation, and SOM can determine whether fuzzy synthetic evaluation is better suited than the Carlson index for Taiwan.

2. Methodology

The Organization of Economic Cooperation Development (OECD) has investigated the water quality of 80 lakes in the world in order to set a standard for determining eutrophic status (see Table 1). In this research, total posphorous (TP), chlorophyll a (chl a), and Secchi disk depth (SD) are the classification factors,

Table 1
Trophic state as a function of nutrient levels defined by OECD

Trophic state	TP (μg/L)	chl a (μg/L)	SD (m)
Oligotrophic	<7.9	<2	>4.6
Oligo–meso	8~11	2.1~2.9	4.5~3.8
Mesotrophic	12~27	3.0~6.9	3.7~2.4
Meso–eutro	28~39	7.0~9.9	2.3~1.8
Eutrophic	>40	>10	<1.7

and the membership functions of the three factors which are most applicable to countries, such as Taiwan, in the subtropic zone are developed. Eutrophic, mesotrophic, and oligotrophic standards defined by OECD are unclear; there is a fuzzy zone between the eutrophic, mesotrophic, and oligotrophic states. This is a feature of the problem which can be solved by fuzzy theory. Based on the OECD report, this study defines three membership functions for the three factors (TP, chl a, and SD) that are relative to the water quality index.

2.1. Determination of membership function

The status of the lake is eutrophic if TP is greater than 40 μg/L according to OECD standards. If TP is lower than 8 μg/L, the status of the reservoir is oligotrophic; and if TP is exactly equal to 20 μg/L, the lake is classified as mesotrophic. Several keypoints defining the value of the membership functions of TP are listed below:

$$\mu_{pe}(40) = 1, \quad \mu_{pe}(20) = 0, \quad \mu_{pm}(8) = 0, \quad \mu_{pm}(20) = 1,$$

$$\mu_{pm}(40) = 0, \quad \mu_{po}(8) = 1, \quad \mu_{po}(20) = 0,$$

where p represents phosphorous; e, eutrophic; m, mesotrophic; and o, oligotrophic. After selecting the keypoints, a membership function should be selected to fit these data sets. The half-trapezoid-shaped distribution function is chosen as the TP membership function. The definition of the half-trapezoid-shaped distribution function is listed in Eq. (1). In this study, we add a power k on the second formula to make the graph smoother and closer to the real-world situation. The parameter k must be fixed to improve the similarity between the membership function and real water quality condition. Parameter k is determined according to the OECD report. When TP ranges from 8 to 12 μg/L, it is difficult to classify the water body as either oligotrophic or mesotrophic; also, when TP ranges from 28 to 40 μg/L, it cannot be characterized as mesotrophic or eutrophic. The degree of fuzziness for the two states is assigned to a maximum value of 0.5 when the TP concentration is exactly equal to 10 and 34 μg/L. Thus, parameter k is determined by Eqs. (2) and (5). The

modified TP membership functions are listed below and ploted in Fig. 1:

$$\mu(x) = \begin{cases} 1, & x \leqslant 0, \\ (b-x)/(b-a), & a < x \leqslant b, \\ 0, & x > bm, \end{cases} \quad (1)$$

$$\mu_{pe}(x) = ((x-20)/20)^{k1} \quad \text{if } 20 < x < 40, \quad (2)$$

$$\mu_{pe}(x) = 1 \quad \text{if } 40 < x, \quad (3)$$

$$\mu_{pm}(x) = 1 - \mu_{p1}(x) - \mu_{p3}(x), \quad (4)$$

$$\mu_{po}(x) = ((20-x)/12)^{k2} \quad \text{if } 8 < x < 20, \quad (5)$$

$$\mu_{po}(x) - 1 \quad \text{if } x < 8, \quad (6)$$

$$x = 34, \quad \mu_{pm}(x) = \mu_{pe}(x) = 0.5, \quad k_1 = 1.943, \quad (7)$$

$$x = 10, \quad \mu_{pm}(x) = \mu_{po}(x) = 0.5, \quad k_2 = 3.802, \quad (8)$$

where p represents phosphorous; e, eutrophic; m, mesotrophic; and o, oligotrophic.

The membership functions for chl a and SD are similarly given as

Membership function of chl a:

$$\mu_{ce}(x) = ((x-5)/5)^{1.943} \quad \text{if } 5 < x < 10, \quad (9)$$

$$\mu_{ce}(x) = 1 \quad \text{if } 10 < x \quad (10)$$

$$\mu_{cm}(x) = 1 - \mu_{ce}(x) - \mu_{co}(x), \quad (11)$$

$$\mu_{co}(x) = ((5-x)/12)^{3.802} \quad \text{if } 8 < x < 20, \quad (12)$$

$$\mu_{co}(x) = 1 \quad \text{if } x < 8, \quad (13)$$

where c represents chl a.

Membership function of SD:

$$\mu_{sde}(x) = ((3-x)/1.3)^{2.642} \quad \text{if } 1.7 < x < 3.0, \quad (14)$$

$$\mu_{sde}(x) = 1 \quad \text{if } 1.7 < x, \quad (15)$$

$$\mu_{sdm}(x) = 1 - \mu_{sde}(x) - \mu_{sdo}(x), \quad (16)$$

$$\mu_{sdo}(x) = ((x-3)/1.5)^{2.235} \quad \text{if } 3.0 < x < 4.5 \quad (17)$$

$$\mu_{sdo}(x) = 1 \quad \text{if } x < 4.5, \quad (18)$$

where sd represents SD.

A final fuzzy membership matrix used for fuzzy synthetic evaluation is derived from synthesizing the trophic membership functions of the three factors (TP, chl a, and SD). The matrix can be formulated as follows:

$$R = \begin{bmatrix} \mu_{pe} & \mu_{pm} & \mu_{po} \\ \mu_{ce} & \mu_{cm} & \mu_{co} \\ \mu_{sde} & \mu_{sdm} & \mu_{sdo} \end{bmatrix}, \quad (19)$$

where p represents total phosphorous; c, chl a; sd, Secchi disk depth; e, eutrophoic; m, mesotrophic; and o, oligotrophic.

2.2. Determination of weights

Chlorophyll a is the most important factor in the determination of trophic status [9]; total phosphorus is ranked as the second [10]. In this study, therefore, we assume an initial judgment matrix J for these factors which is listed below:

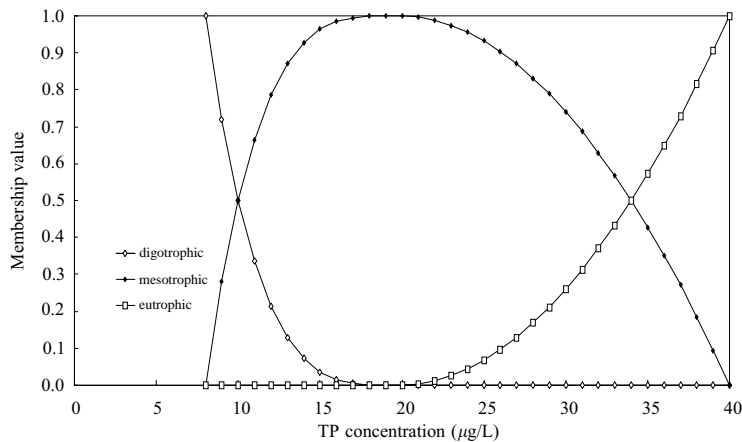$$J = \begin{bmatrix} 1 & \frac{1}{2} & 2 \\ 2 & 1 & 3 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix}. \quad (20)$$

Fig. 1. Membership function for TP.

A normalized matrix D is determined by normalizing the initial matrix J column-by-column. Then, the final weighted vector W can be derived from summing the elements of each row of matrix D and normalizing again to this vector:

$$D = \begin{bmatrix} 0.2857 & 0.2727 & 0.3333 \\ 0.5714 & 0.5455 & 0.5 \\ 0.1429 & 0.1818 & 0.1667 \end{bmatrix}, \tag{21}$$

$$W = \begin{bmatrix} w_p w_c w_{sd} \end{bmatrix} = [0.2972\ 0.5390\ 0.1638]. \tag{22}$$

2.3. Fuzzy synthetic evaluation

In this study, the evaluation set $U = \{$oligotrophic, mesotrophic, eutrophic$\}$ contains three levels, and the factor set $V = \{$chl a, TP, SD$\}$ has three factors. The evaluation matrix of three factors is $R = (u_{ij})_{nxm}$; the weights set is $W = \{w_p, w_c, w_e\}$. When W and R are given, fuzzy synthetic evaluation can be performed as follows:

$$B = W \times R = [w_p w_c w_{sd}] \times \begin{bmatrix} \mu_{pe} & \mu_{pm} & \mu_{po} \\ \mu_{ce} & \mu_{cm} & \mu_{co} \\ \mu_{sde} & \mu_{sdm} & \mu_{sdo} \end{bmatrix}$$
$$= [b_o b_m b_e]. \tag{23}$$

In this study, we use first-level fuzzy synthetic evaluation. The trophic status assessment is determined by the maximum value of the three values (b_o, b_m, b_e). The results of fuzzy synthetic evaluation is summed up with a eutrophication index (EI). The EI value is calculated by

Eq. (24) and can be plotted versus time to reveal the trend of water quality:

$$EI = 1 \times b_o + 2 \times b_m + 3 \times b. \tag{24}$$

3. SOM model development

A sketch of the SOM model is presented in Fig. 2. The SOM consists of two layers—the input layer and the output layer. The neurons of the two layers are fully connected. In addition, each neuron of the output layer is connected laterally to the indicated lateral inhibition. The input pattern is received through the neurons in the input layer. Usually, in the SOM application, it is important to normalize the input vector to unit length:

$$x = X/|X|, \tag{25}$$

where X is the input vector.

Similarly, the synaptic strengths must also be normalized as input vectors:

$$w_i = W_i/|W_i|, \tag{26}$$

where W_i is the synaptic weight vector for ith output neuron.

The dot product between the input vector and the synaptic weights of an output neuron gives the output of that neuron:

$$O_i = w_i \cdot x = \cos è, \tag{27}$$

where O_i is the output of neuron and θ is the angle between the synaptic weights and input vector.

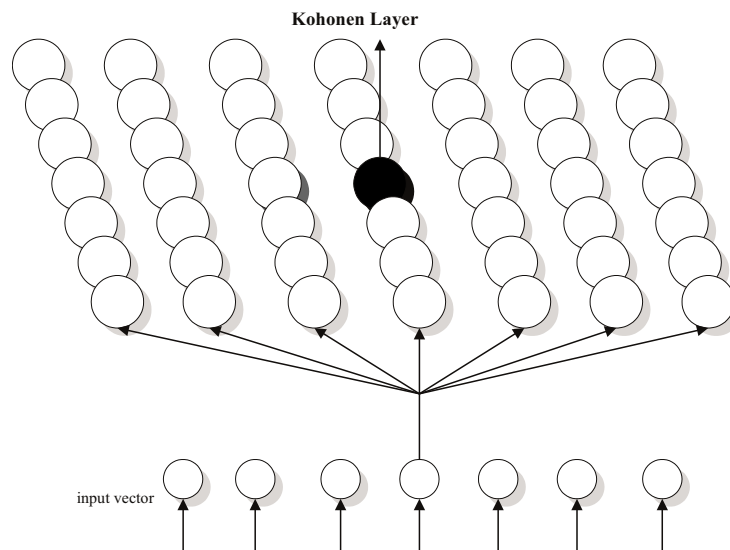The output is simply the *cosine* of the angle between the synaptic weights and input vector. Therefore, O_i

Fig. 2. Structure of the SOM model.

varies from –1 to 1, θ indicating that the angle varies between $0°$ and $180°$. The value of O_i reflects the parallelism between the input vector and the stored synaptic weights for the output neuron. The output is calculated for all the ouput neurons, and the O_i having the highest value is the ''winner''. The training is based on a ''winner takes all'' strategy [7]. Thus, the synaptic weights corresponding to the output neuron with maximum output are updated:

$$w_{i+1} = w_i + \alpha(x - w_i), \tag{28}$$

where α is the learning rate.

The value of α varies from 0 to 1, and it controls the rate of learning. An α of 1 means it learns a new example as soon as it is presented. However, it forgets all previous examples of that class. Similarly, an α of 0 means that the network does not learn at all, but classifies new examples based on previous experiences only.

The number of input units depends on the required resolution of the input pattern. The number of output units is determined by the desired number of classes. In this study, two kinds of SOM models were developed, one with 10×10 neurons on the output layer, and another with 20×20 neurons on the output layer. Once the structure of the SOM model was established, the training process may begin. At first, the SOM model can be initialized to any value of synaptic weight. Initially, the value of α can be set close to 1 and can be reduced gradually as the network progresses. In this study the value of á was set as 0.6 during the training process. After a satisfactory training, the value of á can be kept at 0. Whether or not an SOM model is available for use as a water quality classifier, it is determined by an index, called the average minimum of distance (AMD). The AMD was defined as the average distance between all input vectors and the synaptic weights. As the AMD converged and the AMD value is low enough (< 5), the SOM model can be regarded as a well-trained network.

4. Case study

A case study of SOM and fuzzy synthetic evaluation applied to assess the trophic status of a water body has been implemented at Fei-Tsui Reservoir, situated in northern Taiwan. Fei-Tsui Reservoir (surface = 10.24 km^2; mean depth = 39.68 m; hydraulic retention time = 150.8 days) is one of the most detailed, surveyed reservoirs in Taiwan. The historical database was collected from the management agency of Fei-Tsui Reservoir from 1987 to 1995. There are a total of 110 data records. The measured data were surveyed at different water depths. For the comparison reason, we want to know the change trend of water quality on the same water depth, only data from the top level of the water body were used in this study. There are a total of 18 measured items, such as water temperature, pH, dissoved oxygen, suspended solid, nitrate, and nitrite. However, some of the 18 variables are not eutrophication factors. The monthly average TP, chl a, and SD are used as eutrophication parameters according to the results of factors analysis.

At the beginning of the analysis, each of 110 water quality records was given a unique record number from 1 to 90 and 1* to 20*. The data records were divided into two parts: Record nos. 1–90, which were used to train the SOM model, and record nos. 1*–20*, which were used to test the well-trained SOM model. Each record in the training set was applied iteratively to the 10×10 SOM model to reach a convergence of AMD. When the convergence occurred, training of the SOM model to act as a water quality classifier for this study was completed (Fig. 3). Fig. 3 shows that the AMD converged to a value of 1 after finishing 75 training cycles, indicating that the training phase was completed. Output from the SOM model can be plotted on a two-dimensional map (10×10) to show valuable information. For comparative purposes, the SOM output map was scaled by 0–9 on the x- and y-axis so that each node on the map could be addressed by the value of (x, y).

5. Results and discussion

5.1. Classification results of the 10×10 SOM model

A scattered plot of the 110 records applied to the 10×10 SOM model is illustrated in Fig. 4. Most of the
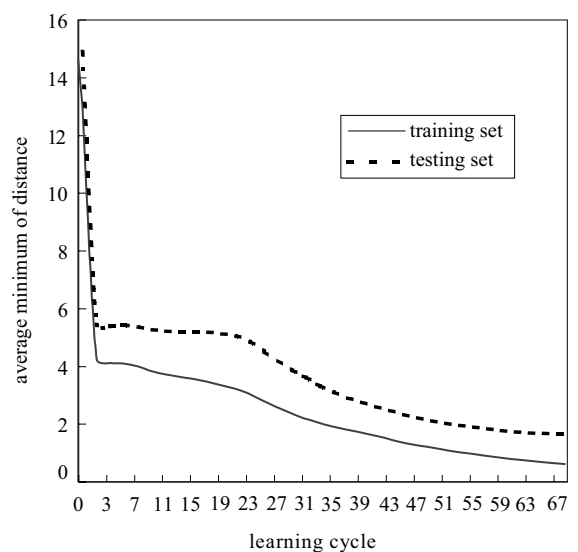

Fig. 3. The convergence of the AMD for the training of the SOM model.

records are situated along the diagonal axis from the lower-left corner to the upper-right corner, and a total of 70 points (63.6%) are summed up. The position mapped by the most records is neuron (0,0), which is labeled as symbol C in Fig. 5. Neuron C includes 21 records

	0	1	2	3	4	5	6	7	8	9
9	5		2						3	10
8	1		1						1	3
7	4	1		1	1			2		
6	2						1			
5	2	1			2	1	1	2		
4	3			1	3					
3			2	2						
2		3	4	1		1				
1	2	3				1			1	
0	21	2		1	3	2	2	1	1	3

Fig. 4. Output map of the 10×10 SOM model.

	0	1	2	3	4	5	6	7	8	9
9	A		2,15						12,14 2*	D
8	63		6*						8	3,23, 7*
7	B	60		90	89				33, 36	
6	17 74						86			
5	7 35	24			29,40	34	76	50 11*		
4	4 ,85 87			42	37,3 , 13*					
3			19* 27	46, 14*						
2		6, 45, 49	77,81, 41,72	55		75				
1	80, 15*	54 ,69 20*			12*				38	
0	C	18,31		53	51,58, 64	68,61	67, 88	66	52	11,26 ,65

Oligotrophic zone   Mesotrophic zone   Eutrophic zone

Tp. is higher or chl. a is lower   chl. a is higher

Fig. 5. Output map of the 10×10 SOM model labeled with corresponding record numbers. (Symbol A includes records 25, 48, 62, 3*, 4*; symbol B, 47, 1*, 5*, 18*; symbol C, 5, 21, 22, 28, 30, 32, 43, 44, 56, 57, 59, 70, 71, 73, 78, 79, 82, 83, 84, 16*, 17*; symbol D, 1, 9, 10, 13, 16, 19, 20, 8*, 9*, 10*).

comprising 19 training cases and two testing cases. The position mapped by the second most records is neuron (9,9), which is labeled as symbol D in Fig. 5. Neuron D includes 10 records comprising seven training cases and three testing cases. Besides neurons C and D, neuron A was mapped by five records; neuron B, four; and neuron (2,2), also four. Other points mapped by 1, 2 and 3 records can also be found in the Fig. 5 but is not described in detail here.

5.2. Identification zone of the oligotrophic state

When all data records mapped to the lower-left corner of the SOM map are studied (Fig. 5), it is observed that they are classified as oligotrophic by both the Carlson index and fuzzy synthetic evaluation (Table 2). Fig. 5 shows that position (0,0) is mapped by records 5, 21, 22, 28, 30, 32, 43, 44, 56, 57, 59, 70, 71, 73, 78, 79, 82, 83, 84, 16*, and 17* (* meaning it belongs to the testing set), all of which are judged as oligotrophic by both the Carlson index and fuzzy synthetic evaluation. The other points along with the sequence (0,1), (1,1), (1,0), and (1,2) are mapped totally by 10 records and they were also classified as oligotrophic by the two methods. Based on these results of identification, the lower-left corner is regarded as the oligotrophic identification zone.

5.3. Identification zone of the mesotrophic state

Several points cluster in the middle of the diagonal of the SOM map (Fig. 5). All of these points are assessed by cross validation using the Carlson index and fuzzy synthetic evaluation (Table 2). The results reveal that they are all mesotrophic. Thus, this area is classified as the mesotrophic identification zone.

5.4. Identification zone of the eutrophic state

Position (9,9) on the SOM map is mapped by record no. 13. This record represents the water quality measurement for October 1988 (Table 2 and Fig. 5) and is judged as eutrophic by both the Carlson index and fuzzy synthetic evaluation. Another position (8,9) on the SOM map is mapped by three records (12, 14, 2*). Record 14 represents the data taken in November 1988; record 2*, in February 1987; and record 12, in September 1988. Records 14, 12, and 2* are neighbors on the SOM map, and record 14 is classified as eutrophic by both the Carlson index and fuzzy synthetic evaluation. Although record 12 is considered to be mesotrophic, examining the Carlson index value of record 12, 49.46 is relatively high. In fact, the water quality is almost eutrophic but it is classified as mesotrophic only because it is not above the critical value of 50; record 2* also encounters the same condition. This is because the
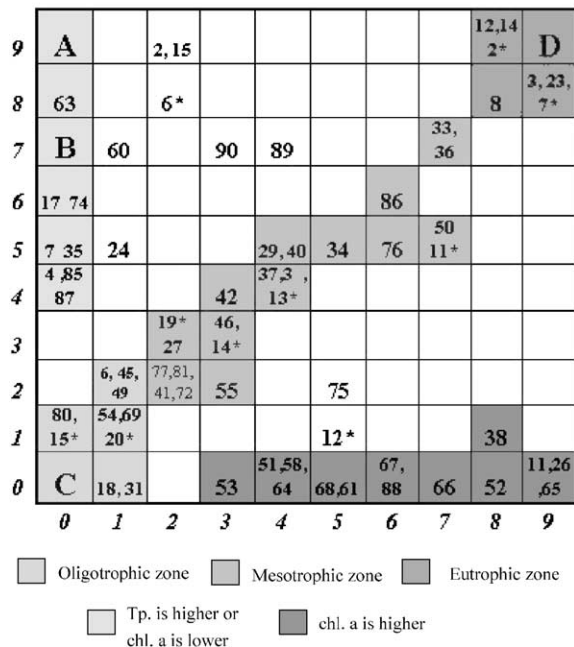
Table 2
Comparison of Carlson index with fuzzy synthetic evaluation

Case no.	Year	SD	TP	chl *a*	Carlson index		Fuzzy synthetic evaluation	
					TSI	Rank	EI	Rank
22	Jun 89	3.10	1.40	3.40	32.54	Oligotrophic	1.65	Oligotrophic
30	Jan 90	5.20	5.20	1.88	34.30	Oligotrophic	1.00	Oligotrophic
43	Jan 91	3.30	6.30	3.19	39.12	Oligotrophic	1.62	Oligotrophic
44	Feb 91	3.20	5.40	3.14	38.49	Oligotrophic	1.61	Oligotrophic
82	Jan 94	5.40	4.47	3.05	34.99	Oligotrophic	1.43	Oligotrophic
84	Mar 94	4.90	4.64	2.56	35.06	Oligotrophic	1.29	Oligotrophic
18	Feb 89	4.40	8.00	4.37	39.89	Oligotrophic	1.56	Oligotrophic
80	Dec 93	4.10	8.20	3.57	39.69	Oligotrophic	1.61	Oligotrophic
72	Apr 93	3.60	8.90	4.04	41.10	Mesotrophic	1.75	Mesotrophic
19*	May 95	2.59	8.80	4.27	42.27	Mesotrophic	1.77	Mesotrophic
37	Aug 90	2.80	11.10	4.88	43.97	Mesotrophic	1.91	Mesotrophic
40	Nov 90	2.10	11.70	4.54	45.37	Mesotrophic	1.99	Mesotrophic
33	Apr 90	2.70	14.00	5.21	45.45	Mesotrophic	1.98	Mesotrophic
34	May 90	2.50	12.40	4.65	44.88	Mesotrophic	1.96	Mesotrophic
13*	Dec 94	4.60	12.00	5.59	42.40	Mesotrophic	1.78	Mesotrophic
76	Aug 93	2.10	13.30	5.39	46.53	Mesotrophic	2.03	Mesotrophic
39	Oct 90	3.00	11.30	5.57	44.16	Mesotrophic	1.92	Mesotrophic
46	Apr 91	3.20	9.40	4.90	42.56	Mesotrophic	1.81	Mesotrophic
23	Jul 89	2.30	11.10	3.99	48.48	Mesotrophic	2.07	Mesotrophic
20	Apr 89	2.50	22.60	5.46	48.24	Mesotrophic	2.02	Mesotrophic
12	Sept 88	2.80	29.90	6.27	49.46	Mesotrophic	2.11	Mesotrophic
13	Oct 88	2.40	25.20	7.99	50.14	Eutrophic	2.20	Eutrophic
14	Nov 88	0.90	30.80	5.75	54.78	Eutrophic	2.27	Eutrophic

Note: * represents data belonging to the testing set; SD represents Secchi disk depth (m); TP, total phosphorous (μg/L); chl *a*, chlorophyll a (μg/L).

Carlson index is deficient in the identification of continuous information. Record 12 is the best representative for this condition. The same results are also found in records 1, 9, 10, 13, 16, 20, 23, 26, 33, and 36. These records are all located in the upper-right corner of the SOM map and their Carlson index values are above 45 and close to the critical value of 50. However, the SOM model can resolve this problem by mapping the data as close to eutrophication zone as it can. Thus, according to these results of identification, the upper-right corner of the SOM output map is regarded as the eutrophic identification zone.

5.5. Conflicting assessment results

Fig. 5 shows that record 65 is situated on position (9, 0), which is the lowest right cell. From Table 2, it can been seen that record 65 represents the data taken in October 1992. This record is classified as mesotrophic by the Carlson index but as eutrophic by fuzzy synthetic evaluation. Besides record 65, there are four records clustered around the lower-right corner of the SOM output map: they are records 26, 52, 66, and 67. Especially for records 66 and 67, taken in November and December 1992. Again, all of these four records

were classified as mesotrohic by the Carlson index, but as the eutrophic by fuzzy synthetic evaluation.

Those data situated on the diagonal and lower-right corner of Fig. 5 have been discussed. However, there are some data, situated on the upper-left corner of the SOM output, which have not been discussed yet. Record 17, situated at position (0, 6), was taken in January 1989. This record was classified as oligotrohic by the Carlson index, but as mesorophic by fuzzy synthetic evaluation (Table 3). Besides record 17, record 74 is also situated in the upper-left corner on the SOM output map, at position (0, 6). It is also classified as oligotrohic by the Carlson index but as mesotrophic by fuzzy synthetic evaluation.

As shown in Table 3, records 2, 15, 3*, and 4* all were classified as mesotrophic by the Carlson index but oligotrophic by fuzzy synthetic evaluation. These data can be found within the same area on the SOM output map (Fig. 5). Although records 2, 15, 3*, and 4* have conflicting assessment results, records 2 and 15 are not exactly the same as records 17 and 74. Therefore, SOM allocated different data storage areas for these four records: 2, 15, 3*, and 4*.

The conflicting data shown on the SOM output map can be explained by some investigations performed in

Table 3
Conflicting assessment results conducted by Carlson index and Fuzzy synthetic evaluation

Record no.	Year	SD	TP	chl a	Carlson index		Fuzzy synthetic evaluation	
					TSI	Rank	EI	Rank
26	Oct 89	3.00	13.50	14.47	48.12	Mesotrophic	2.51	Eutrophic
52	Oct 91	2.80	9.10	14.39	46.57	Mesotrophic	2.33	Eutrophic
65	Oct 92	3.10	8.3	19.83	46.70	Mesotrophic	2.27	Eutrophic
66	Nov 92	4.00	6.50	13.14	42.97	Mesotrophic	2.18	Eutrophic
67	Dec 92	4.40	5.00	11.69	40.90	Mesotrophic	2.10	Eutrophic
3*	Mar 87	2.90	14.70	1.29	40.78	Mesotrophic	1.45	Oligotrophic
4*	Apr 87	3.00	14.20	1.88	41.68	Mesotrophic	1.44	Oligotrophic
17	Jan 89	5.20	12.60	2.47	39.37	Oligotrophic	1.51	Mesotrophic
74	Jun 93	5.96	11.30	2.90	38.72	Oligotrophic	1.61	Mesotrophic
2	Dec 87	2.00	17.10	1.71	44.20	Mesotrophic	1.54	Oligotrophic
15	Dec 88	2.00	17.10	1.71	44.20	Mesotrophic	1.54	Oligotrophic

Note: * represents data belonging to the testing set; SD represents Secchi disk depth (m); TP, total phosphorous (μg/L); chl a, chlorophyll a (μg/L).

Taiwan. A research report [11] reveals that a reservoir water body does not stabilize until the nutrients that existed before the reservoir began operation are exhausted. At least 5 years are needed before the water quality becomes more stable [11]. This phenomenon is also proven by our previous investigation [5] using fuzzy synthetic evaluation. Eutrophication of Fei-Tsui Reservoir caused by overturn ceased after 1992, which is exactly 5 years from the first year of operation (1987). However, the Carlson index is unable to reveal the same information, which is why most of the conflicting data fell in 1988–1992.

5.6. Classification results of the 20 × 20 SOM model

When 110 records were mapped to the 10 × 10 SOM output layer, it was easy to find that some records fell into the same neuron. Thus, another SOM model, one with 400 output nodes (20 × 20), was designed to analyze the same data set. The output map showed that the classification results are approximately similar to that of the 10 × 10 SOM model, but the direction of the diagnostic axis changed from lower-left to upper-right in the 10 × 10 SOM model to upper-left to lower-right 20 × 20 SOM model. There are still three identification zones on the axis: the upper-left corner of the axis represents the oligotrophic zone; the middle of the axis, the mesotrophic zone; the lower-right of the axis, the eutrophic zone. For the 20 × 20 SOM, the number of data records mapped into the same neuron is six, less than that of the 10 × 10 SOM. This means that the resolution of classification is determined by the number of output neurons. The more the neurons used in the SOM output layer, the stronger the ability of SOM model to identify the diversity between data.
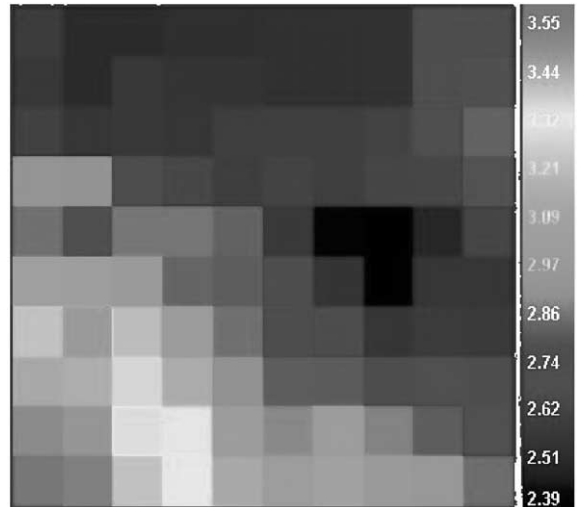

Fig. 6. Reference vector component plane map of Secchi depth for 1987–1995.

5.7. Vector component plane map of the SOM output

In this study, because the input vector is composed of three factors, SD, TP, and chl a, the SOM output map consists of three component planes: SD, TP. and chl a (Figs. 6–8). Figs. 6 and 7 show that the previously mentioned diagnostic axis of water quality is formed according to SD and TP. The identification zone smoothly changes from the lower-left corner to the upper-right corner. However, the trend of change of chl a is from the upper-left corner to the lower-right corner on the component plane map of the chl a (Fig. 8). In Fig. 5, those data that fall in the bottom of the SOM
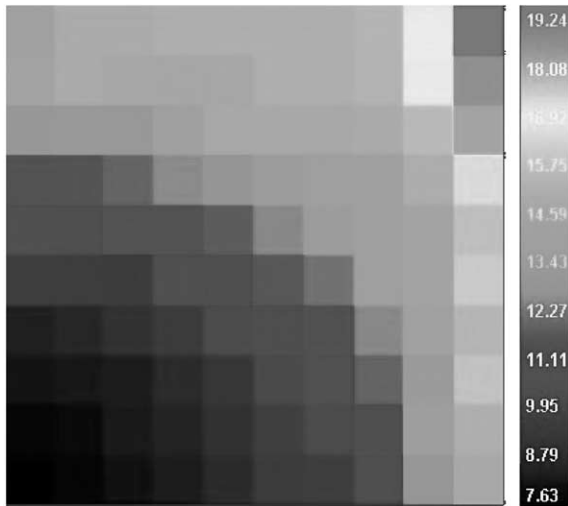
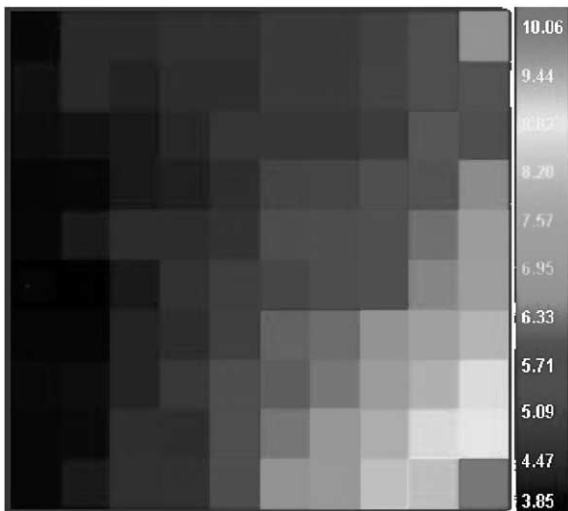Fig. 7. Reference vector component plane map of total phosphorus for 1987–1995.

Fig. 8. Reference vector component plane map of chlorophyll *a* for 1987–1995.

output map are summarized in Table 3. Records 75 and 12* are slightly higher in chl *a*. Along the *x*-axis, the concentrations of chl *a* for records 53, 51, 61, 88, 28, and 52–11 gradually increase (Table 3). Besides, along the *y*-axis of Fig. 5, especially in neuron $(0, 9)$ and $(0, 8)$, records 25, 48, 62, and 63 all have lower chl *a* concentrations (Table 4). Thus, whether the water quality is abnormal or not, it mainly depends on the chl *a* concentration (Fig. 8).

5.8. Application potential of the SOM

Generally, Carlson index is good for evaluating the water quality on site and for real-time purpose.

However, for long-term water quality investigation, Carson index may lose something important that are hidden behind those records, because it does not deal with the differences among data. For this reason, a more sensitive local index should be developed for each area. Thus, Fuzzy synthetic evaluation was proposed to evaluate reservoir water quality in Taiwan. If Fuzzy synthetic evaluation needs to be used in other countries, some parameters should be changed. Similar to the development process of Fuzzy synthetic evaluation, when a new local trophic index is proposed, the first thing to do is to validate whether the local trophic index is good or not. The advantage of the SOM model is an impersonal feature; that is, SOM with the ability to learn and organizing information without being given associated-dependant output values for the input patterns. In this study, SOM was proposed to be used to validate whether the local water quality index is good or not and the results of the SOM showed that it is also a complementary diagnostic tool. For example, if some data records are evaluated to give inconsistent results by the Carlson index and Fuzzy synthetic evaluation, we can double check the SOM result to identify the final results.

6. Conclusion

In this research, one complementary classification model, the self-organizing map (SOM), for diagnosing water quality has been developed and illustrated with a case study of trophic status assessment for Fei-Tsui Reservoir in Taiwan. In our previous investigation, fuzzy synthetic evaluation identified eutrophication in the reservoir caused by a short-term overturn phenomenon in 1989, 1991, and 1992, but the Carlson index did not. However, in this study, the inconformity of evaluation between the Carlson index and Fuzzy synthetic evaluation can be identified by the SOM model. All of the inconsistent records are mapped in the abnormal zone, which are located in the upper-left and lower-right corners of the 10×10 SOM map.

Besides two abnormal zones, three main clusters can be observed on the SOM output map; the first one, situated in the lower-left corner of the diagonal is the identification zone of oligotrophic status; the second, situated in the middle of the diagonal is the identification zone of mestrophic status; and the third, situated in the upper-right of the diagonal is the identification zone of eutrophic status. Based on the results of the 10×10 and 20×20 SOM models, the resolution of classification is determined by the number of output neurons. The more the neurons used in SOM output layer, the stronger the ability of the SOM model to identify the diversity between data. In addition, based on the analysis of the component planes of the SOM output

Table 4
Special water quality records found in SOM output

Case no.	Year	SD	TP	chl a	Special reason
Total average		3.29	12.23	4.92	
75	Jul 93	2.10	9.90	5.53	chl a is slightly higher
12*	Nov 94	3.70	9.60	5.90	chl a is slightly higher
53	Nov 91	3.50	8.40	5.07	chl a is slightly higher
51	Sep 91	2.70	7.40	5.77	chl a is slightly higher
61	Jun 92	2.40	8.60	6.37	chl a is slightly higher
88	Jun 94	2.80	8.00	8.40	chl a is heavily higher
38	Nov 90	1.70	13.70	8.64	chl a is heavily higher
52	Oct 91	2.80	9.10	14.39	chl a is heavily higher
11	Aug 88	2.70	41.40	43.59	chl a is heavily higher
89	Aug 94	2.80	14.60	4.40	chl a is slightly lower
90	Sept 94	1.90	13.60	4.10	chl a is slightly lower
25	Sept 89	3.00	14.90	2.11	chl a is heavily lower
48	Jun 91	2.10	13.60	3.20	chl a is heavily lower
62	Jun 92	2.30	14.30	3.55	chl a is heavily lower
63	Aug 92	2.60	14.20	3.60	chl a is heavily lower

Note: * represents data belonging to the testing set; SD represents Secchi disk depth (m); TP, total phosphorous (μg/L); chl a, chlorophyll a (μg/L).

map, a diagnostic axis of water quality is formed according to SD and TP. Whether the water quality is abnormal or not, it mainly depends on the concentration of chl a. For the future SOM application, as long as the SOM model is well-trained, new records can be assessed and classified as either one of three trophic levels or special cases. If special water quality conditions are expressed on the SOM output, these data can show whether total phosphorus or chlorophyll a is higher than usual.

References

[1] Carlson R. A trophic state index for lakes. Limnol Oceanogr 1977;22(2):361–9.

[2] Morihiro A, Outoski A, Kawai T, Hosome M, Muraoka K. Application of modified Carlson's trophic state index to Japanese lakes and its relationship to other parameters related to trophic state. Res Report Natl Inst Environ Stud 1981;23:12–30.

[3] Weiss CM, Francisco DE, Campbell PH. Water quality study. A report to the Wilmington District, US Army Corps of Engineer, Department of Environmental Science and Engineering, University of North Carolina at Chapel Hill, 1985.

[4] Kung H, Ying L, Liu YC. A complementary tool to water quality index: fuzzy clustering analysis. Water Resources Bull 1992;28(3):525–33.

[5] Lu R-S, Lo S-L, Hu J-Y. Analysis of reservoir water quality using fuzzy synthetic evaluation. Stochastic Environ Res Risk Assessment 1999;13(5):327–36.

[6] Kohonen T, OJA E, Simula O, Visa A, Kangas J. Engineering applications of the self-organizing map. Proc of the IEEE 1996;84(10):1358–84.

[7] Mukherjee A. Self-organizing neural network for identification of natural modes. J Comput Civil Eng 1997;11(1):74–7.

[8] Kohonen T. Self-organizing maps. New York, Inc: Springer, 1995.

[9] Minerva NM. Evaluation of nutrient-chlorophyll relationships in the Rybinsk Reservoir. Water Sci Technol 1993;28(6):25–8.

[10] Nedoma J. Phosphorus deficiency diagnostics in the eutrophic Rimov Reservoir. Water Sci Technol 1993;28(6):75–84.

[11] Chang SP, Wen CK. Effect of nutrients on the eutrophication of a reservoir in the initial stage of operation. Proceedings of the Eighth Annual Conference on Environmental Planning and Management. Tainan, Taiwan, November 18–19, 1995. p. 36–45 (in Chinese).