Dynamic risk assessment model for water quality on projection pursuit cluster

Jun Zhao, Juliang Jin, Xiaomin Zhang and Yaqian Chen

ABSTRACT

With the aim of reducing the losses from water pollution, a dynamic risk assessment model for water quality is studied in this paper. This model is built on the projection pursuit cluster principle and risk indexes in the complex system, proceeding from the whole structure and its component parts. In this paper, the fuzzy analytic hierarchy process is used to screen out index system and determine index weight, while the further value of an index is simulated by hydrological model. The proposed model adopts the comprehensive dynamic evaluation method to analyze the time dimension data, and evaluates the development tendency by combining qualitative analysis with quantitative analysis. The projection pursuit theory is also employed for clustering the spatial dimension data, the optimal projection vector for calculating risk cluster type to compartmentalize risk, and then local conditions for proposing the regulation scheme. The applicational results show that the model has the strong logic superiority and regional adaptability with strict theoretical system, flexible methods, correct and reasonable results and simple implementation to provide a new way for research on risk assessment models of water quality.

Key words | dynamic risk assessment, fuzzy analytic hierarchy process, hydrological model, projection pursuit cluster, water quality

Jun Zhao (corresponding author)
Applied Hydrometeorological Research Institute,
Nanjing University of Information Science & Technology,
Nanjing, 210044,
China
and
College of Hydrology and Water Resources,
Hohai University,
Nanjing, 210098,
China
E-mail: *zsmzyq@126.com*; *zhangxm@ncpe.com*

Juliang Jin
School of Civil Engineering,
Hefei University of Technology,
Hefei 230009,
China

Xiaomin Zhang
North China Power Engineering Co. Ltd, of China Power Engineering Consulting Group,
Beijing 100120,
China

Yaqian Chen
Pearl Water Resources Protection Institute,
Guangzhou 510611,
China

INTRODUCTION

Water is the source of life, a necessity of production, and fundamental for ecology. Water resource has been an irreplaceable natural resource to human beings for life and an indispensable economic resource to society for development. According to incomplete statistics, almost 95% of industrial waste is discharged into rivers and lakes without treatment in China. Many bodies of water are suffering from point pollution and non-point pollution to different degrees in an aggravating trend. Water shortage and water pollution are increasingly serious year after year. From the perspective of system theory, the determination of risk indexes in water quality systems, the analysis of the current and future situation of water quality, the prediction of risk level and establishment of the corresponding regulation countermeasures beforehand are the important parts for setting up risk assessment systems for water quality suited to China's national conditions. Recently, a comprehensive risk assessment of water quality has become a hot topic in the global hydrological field (Chen *et al.* 2007; Gao *et al.* 2007; Yao & Zhang 2009; Zhang & Dong 2009; Wang *et al.* 2010a).

There are many risk assessment methods, referring to vector graphics, mathematical statistics and evidence, fuzzy mathematics, grey correlation, evolutionary modeling and other theories (He 2007; Zhao 2011). Yang *et al.* (2004) used a genetic projection pursuit interpolation model to comprehensively evaluate water quality, solved the problem of incompatible results of all the single indexes, and raised the model precision for water quality assessment. Jin *et al.* (2007b) applied the entropy coupling method of correspondence factor analysis and projection pursuit to evaluate Chaohu lake water quality security, unearthed the various subjective and objective information in complex system, and

doi: 10.2166/nh.2012.095

enhanced the reliability of comprehensive evaluation results. Based on the ideal solution and grey relational analytical method by means of projection pursuit method, Wang *et al.* (2008) established the combinatorial evaluation and optimization model to dispose of non-linear problems, while the index weights are objective, and confirmed the combinational preference coefficients and the identification coefficients of grey correlation degree scientifically. Wang & Li (2007) proposed the dynamic cluster model on projection pursuit principle to analyze climate zoning, which opened a new way for solving the multi-factor cluster analysis problem. Li *et al.* (2007) made use of the triangular fuzzy numbers theory to river water quality simulation and risk assessment research, constructed a risk fuzzy assessment model for river water quality, and analyzed the risk condition of river water quality in a specified section. Li & Shi (2007) discussed the imperfection and fuzziness indexes of water quality risk, and constructed the fuzzy probabilistic model for calculating risk rate on water quality. Yao & Ni (2009) presented a projection pursuit classification model based on K-Means dynamic cluster to comprehensively evaluate multi-factor's influence, overcame the uncertain cutoff radius problem, and obtained the dynamic cluster results by actual data without artificial judgment. Wang *et al.* (2010b) comprehensively evaluated river health grade by the entropy projection pursuit model and took the maximum entropy value as an objective function to make the projection value carry the evaluation index system's variant information as far as possible.

The water quality system is a complex system subject to the dual effects from nature and human activities, and its risk assessment still lacks the unified evaluation index system and the operable quantitative evaluation model, the problems of which are described from three points of view as follows (Wu *et al.* 2009).

Firstly, according to incommensurability and incompatibility among evaluation indexes, is how to screen risk evaluation index system and determine index weight reasonably. Secondly, taking into account the characteristics of numerous risk evaluation indexes with the complicated non-linear relationship, is how to establish an effective prediction model of risk evaluation index on the study area. Thirdly, in view of inconsistent information from between point value of evaluation index samples and interval values of evaluation criterion grades, is how to create the functional relationship between sample values and criterion grades. However, it is usually difficult for the conventional methods to combine with or make use of experts' experiences during the risk assessment process. It is hard to carry out the meta-synthesis of qualitative analysis and quantitative calculation. Most, seldom consider the risk assessment concept nor build a coupling system, let alone give discussion on macroscopic and microscopic scale nesting.

This paper studies the dynamic risk assessment model for water quality on the projection pursuit cluster principle. It integrates the projection pursuit cluster method with the comprehensive dynamic evaluation theory, puts the screening, prediction and evaluation sub-models together, analyses water quality risk zoning for the first time, and explores the study work of the assessment model under the risk from the whole structure and its component parts, which is also verified through actual example in Taihu basin. Specifically, firstly it uses the fuzzy analytic hierarchy process to screen risk assessment index system and determines index weights. Secondly, it applies a hydrological model to make water quality simulation to get index values in the plain area of the Taihu basin. Thirdly, it discusses the development tendency of water quality in the whole area with the dynamic risk evaluation model for water quality on the projection pursuit cluster principle, the purpose of which is to compartmentalize risk level so as to conduct risk assessment research, hereby establishing regulation measures and implementing emergency plans.

Dynamic risk assessment model for water quality

The dynamic risk assessment model for water quality is composed of index screening, prediction and evaluation sub-models. The screening sub-model, the basis of assessment, gives the fuzzy complementary judgment matrix according to the actual data of study area and experts' experiences, and uses the fuzzy analytic hierarchy process based on an accelerating genetic algorithm to calculate index's weight of matrix in order to establish a risk evaluation index system for water quality. A prediction sub-model, the prerequisite for assessment, mainly selects the appropriate hydrological model with basin concrete conditions to predict each water quality index value in the risk evaluation index system, and thus masters the future development of water quality. An evaluation sub-model, the core part of the technology, firstly employs the projection pursuit cluster method for water quality cluster evaluation, secondly applies the comprehensive dynamic evaluation theory to analyze its tendency, and then obtains a risk level of water quality.

Screening sub-model

With the aim of getting a scientific analysis for water quality, risk indexes should be selected to reflect water quality information as far as possible. However, excessive indexes may bring about a complexity of evaluation process and affect the reliability of evaluation results due to unnecessary data. Based on the principles of integrity, simplicity, hierarchy, and maneuverability, this paper firstly proposes a risk evaluation index system for water quality, secondly provides qualitative analysis with expert's consultation and practical experience, and thereby attempts to provide quantitative analysis. In this paper, risk evaluation index system for water quality $\{x_{jk}|j=1,2,\ldots,m;k=1,2,\ldots,n_j\}$ is screened out via the fuzzy analytic hierarchy process that is based on an accelerating genetic algorithm (Jin *et al.* 2007a). Experts are invited to compare the evaluation importance among the risk indexes to construct the fuzzy complementary judgment matrix $P=(p_{kl})$, satisfying the requirement: $0 \le p_{kl} \le 1$, and $p_{kl}+p_{lk}=1$ $(k,l=1,2,\ldots,n_j)$, where p_{kl} represents the degree of index k superior to index l. When $p_{kl}>0.5$, it shows that experts think index k is more important than index l, and the larger p_{kl} suggests that index k is more important, and vice versa. If P does not have satisfactory consistency, the revision is needed. Assuming that the correction judgment matrix of P is described by $Q=(q_u)$, every index weight of Q is denoted as $\{w_{jk}|j=1,2,\ldots,m;k=1,2,\ldots,n_j\}$, and

$$\begin{aligned}\min \mathrm{CIC}(n_j) &= \sum_{k=1}^{n_j}\sum_{l=1}^{n_j}|q_{kl}-p_{kl}|/n_j^2 \\ &+ \sum_{k=1}^{n_j}\sum_{l=1}^{n_j}\left|0.5(n_j-1)(w_{jk}-w_{jl})+0.5-q_{kl}\right|/n_j^2\end{aligned}$$

$$\begin{aligned}&\text{s.t. } 1-q_{lk}=q_{kl}\in[p_{kl}-d,p_{kl}+d]\cap[0,1] \\ &\quad\times(k=1,2,\ldots,n_j-1;l=k+1,\ldots,n_j)\\ &q_{kk}=0.5(k=1,2,\ldots,n_j)\\ &w_{jk}>0(k=1,2,\ldots,n_j),\ \sum_{k=1}^{n_j}w_{jk}=1 \end{aligned} \tag{1}$$

where $\mathrm{CIC}(n_j)$ is the consistency index coefficient, d is a non-negative value, selected from [0, 0.5] by experts' experience, and the remaining symbols are as described above. The optimal judgment matrix for fuzzy consistency of P is Q that makes formula (1) smallest, the weights and the upper triangular elements of correction judgment matrix are optimizing variables, and the fuzzy complementary judgment matrix P in n_j order of sub-system j totally has $n_j(n_j+1)/2$ independent optimizing variables. Accelerating genetic algorithm is a generally global optimization method simulating the rules of 'Survival of the fittest' and 'Exchange mechanism of chromosome information in community interior'. It is brief and effective to solve the problem of formula (1). When $\mathrm{CIC}(n_j)$ is smaller than a critical value, it can be regarded that P has the satisfactory consistency, and the weights w_{jk} of every index obtained is acceptable. Otherwise, it needs an increase in parameter d or modification of the initial judgment matrix P until it gets the satisfactory consistency. To improve the reliability of index screening, N_e experts are invited to create N_e fuzzy complementary judgment matrixes independently, the fuzzy analytic hierarchy process with accelerating genetic algorithm is employed to obtain N_e weights of evaluation index

$$\{w_{jkl}|j=1,2,\ldots,m;k=1,2,\ldots,n_j;l=1,2,\ldots,N_e\},$$
$$\bar{w}_{jk}=\sum_{l=1}^{N_e}w_{jkl}/N_e(j=1,2,\ldots,m;k=1,2,\ldots,n_j),$$

and then N_j indexes with relatively larger mean value and smaller standard deviation of weights are chosen to form the final risk evaluation index system of water quality $\{x_{jk}|j=1,2,\ldots,m;k=1,2,\ldots,N_j\}$.

Prediction sub-model

There are many different prediction models and selection depends on the specific circumstances of a study area. This paper considers Taihu basin, for example. According to actual conditions, it uses the Taihu basin model, with three parts of database system, scheme management system and model system, to predict risk indexes of water quality. The database system mainly designs the rational and efficient database structure, collects basic data and realizes data management. The scheme management system mainly brings the model base, geographic information system, database technology and other technology together to generate an integral operation platform. The model system mainly includes a rainfall-runoff model, water quantity and water quality model of river networks, wastewater load model, the Taihu lake current model and water quality model of the Taihu lake area (Cheng *et al*. 2006). The actual process of prediction is as follows. Firstly, it divides the plain area of the Taihu basin into 15 hydraulic districts on the basis of terrain features, river network, water resources characteristics and the whole layout of watershed, summarizes the river cross-sections of the Taihu basin by river properties, digitizes information with the classification of basin underlying surface, and makes the river network generalized and geography information unified. Secondly, it selects water quality as the study object, with tide level processes in the Yangtze River and Hangzhou Bay and meteorological data of the basin as the boundary conditions. Thirdly, it assumes the data in January as the initial concentration, takes 15 min for the calculation period and microsegment for calculation space step, and outputs the average concentration for each generalized river every day based on the results of component concentration in sections. Finally, the average concentration of water quality for every year is obtained by the statistical calculation data from early February to the end of December (Zhao 2007).

Evaluation sub-model

The projection pursuit is a statistical method which deals with a multi-factor complicated problem. It offers projection from high-dimensional data to low-dimensional space, and explores the features of high-dimensional data by analyzing projection characteristics of low-dimensional space (Jin *et al*. 2008). Density bandwidth is the only parameter in the projection pursuit model which is usually decided on experience or trial calculation and lacks theoretical basis, so the dynamic cluster principle is introduced into this paper to construct a projection index in order to build up the evaluation sub-model. If the jth index of ith sample is x_{ij}^0, where $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$, n is the number of samples and m is the number of indexes, the procedures of establishing the evaluation sub-model are as follows.

1. Data non-dimensional-normalization. On account of various dimensions of risk evaluation indexes, non-dimensional normalization needs to be conducted for each index data to eliminate the dimension effect and its formula is expressed as:

$$x_{ij} = (x_{ij}^0 - x_{j\min}^0)/(x_{j\max}^0 - x_{j\min}^0) \tag{2}$$

   where, $x_{j\max}^0$ and $x_{j\min}^0$ are the maximum and minimum values of index j in the samples, respectively.
2. Linear projection. The projection is observing data from different angles to seek out the optimal observation angle, which is the best projection direction that reflects data characteristics farthest and excavates data information fully. This method is not only visualized but also convenient for conventional methods to carry on the analysis of high-dimensional data. Linear projection is selected in this paper, which conducts the projection from high-dimensional data to one-dimensional linear space. It is assumed that a is the unit projection direction vector of m dimensional, whose component is $a_1, a_2, \ldots, a_m$, then z_i, one-dimensional projection eigenvalue of x_{ij}, can be expressed as:

$$z_i = \sum_{j=1}^{m} a_j x_{ij} \quad (i = 1, 2, \ldots, n) \tag{3}$$

   where $z = (z_1, z_2, \ldots, z_i, \ldots, z_n)$ is the set of projection eigenvalues.
3. Projection index construction. This step is the key for setting up the evaluation sub-model, the projection principle that ought to be followed from high-dimensional data to low-dimensional space, and the basis of seeking the optimal projection direction. This paper constructs the projection index based on the dynamic cluster principle. Firstly, it is assumed that $s(z_i, z_k)$ is the absolute distance between two random projection eigenvalues and $s(z_i, z_k) = |z_i - z_k|$, $(k = 1, 2, \ldots, n)$; and it divides the cluster samples into N classes, where $2 \le N < n$, and takes $\Theta_h (h = 1, 2, \ldots, N)$ to describe the projection eigenvalues set of the hth classification sample, where

$$\Theta_h = \{z_i | d(A_h - z_i) \le d(A_t - z_i)\}, (t = 1, 2, \ldots, N, t \ne h) \tag{4}$$

In the formula, $d(A_t - z_i) = |z_i - A_t|$, $d(A_h - z_i) = |z_i - A_h|$, A_h and A_t are the initial cluster core of the hth and tth classification respectively, which can be iteratively replaced by the mean value of projection eigenvalue of classified sample in practical operation. Secondly, the neighboring degree of a similar sample can be expressed by a similar centralization degree $d_\mathrm{d}(a)$, whose formula is as follows.

$$d_\mathrm{d}(a) = \sum_{h=1}^{N} d_h(a) \tag{5}$$

In the formula,

$$d_h(a) = \sum_{z_i, z_k \in \Theta_h} s(z_i, z_k),$$

and the smaller $d_\mathrm{d}(a)$, the higher the centralization degree of similar samples. The discrete degree among different samples may be described by different dispersion degree $s_\mathrm{s}(a)$, whose formula is as follows.

$$s_\mathrm{s}(a) = \sum_{z_i, z_k \in Z} s(z_i, z_k) \tag{6}$$

In this formula, the bigger $s_\mathrm{s}(a)$, the higher the discrete degree among different samples. Finally, the projection index constructed by dynamic cluster may be expressed as:

$$Q_Q(a) = s_\mathrm{s}(a) - d_\mathrm{d}(a) \tag{7}$$

It is obvious that the bigger $s_\mathrm{s}(a)$, the longer the distance between different samples, namely the different samples are highly dispersed. Otherwise, the smaller $d_\mathrm{d}(a)$, the shorter the distance among similar samples, namely similar samples are highly concentrated. Therefore, when $Q_Q(a)$ gets the maximum value, the clustering purpose that different samples are dispersed and similar samples are concentrated as far as possible has been realized.

4. Modeling and its optimization. When formula (7) has the maximum value, the optimal projection direction and the clustering results reflecting the data characteristics are obtained. Then the evaluation sub-model can be described as the nonlinear optimization problem shown in formula (8).

$$\begin{cases} \max Q_Q(a) \\ \|(a)\| = 1 \end{cases} \tag{8}$$

This paper uses the accelerating genetic algorithm to solve this problem, the steps of which are outlined as the following.

(i) Randomly generate the unit projection direction vector a of m dimensional for p groups ($p \geq 300$ is suggested), that is to say, create the parent population, and take formula (3) to get the projection eigenvalue vector z for p groups.
(ii) Calculate $s_\mathrm{s}(a)$ and $d_\mathrm{d}(a)$ separately on the basis of z, then get p projection indexes $Q_Q(a)$ by formula (7).
(iii) Make a fitness evaluation with $Q_Q(a)$, and the bigger $Q_Q(a)$ the higher the individual fitness, then gradually generate the first, second and third sub-generation populations through selection, crossover and mutation operations in the genetic algorithm to get a new projection direction vector.
(iv) Separately calculate $Q_Q(a)$ corresponding to the first, second and third sub-generation populations, rank them from great to little, and select the previous p groups as the new projection direction vector based on the principle that the bigger, the more superior, then turn to step (i). If those are fewer than p groups, it should be completed for p groups in the random generation method.
(v) When the difference between fore-and-aft generation projection indexes satisfies the design requirement, the calculation is stopped and the optimal projection direction vector and the final clustering results are obtained.
(vi) Accelerating circulation is put forward to take the parameter change space of excellent individual datum after the first and second iterations as the new change space of model parameter on orthogonal design method. If circulation is continued throughout, the changing space of excellent individual datum is gradually going to contract to approach the optimal point. The whole method

will be over when it achieves the prearranged accelerating number or the optimized principle value of optimal individual datum is less than a pre-specified value.

APPLICATION EXAMPLE

Basin general situation

The plain area of Taihu basin was chosen as an example and is located in the south of the Yangtze River Estuary, between Qiantang River and Hangzhou Bay. It is situated in a subtropical monsoon climate zone with the climate characteristics of four distinct seasons, abundant rainfall, frequent typhoons, etc. The total basin area is 36,895 km^2. The multi-year average precipitation from 1,956 to 2,000 is 1,177 mm, but the precipitation changes greatly in different years and its annual distribution is very uneven. Generally, the flood season from May to September can account for up to 60% of the precipitation for the whole year.

Screening, prediction and evaluation model

The technical thrust of this paper is that firstly, it uses the fuzzy analytic hierarchy process (FAHP) based on the accelerating genetic algorithm (AGA) to establish a risk index evaluation system and determine every index weight. Secondly, on the one hand, using all the 1,483 generalization rivers in the 15 hydraulic districts, it applies the Taihu basin model to perform system coupling for water quality index prediction from 2000 to 2010. Then it adopts the dynamic risk assessment model for water quality based on the projection pursuit cluster principle to compartmentalize risk in order to study risk assessment. On the other hand, using actual data of water quality from 2000 to 2009, it repeats the assessment process mentioned above. Finally, it compares the recursive results via actual data with the results of the screening, prediction and evaluation model from 2000 to 2010, utilizes rational assessment results and local conditions to reduce the possible risk impact by adjusting the strategy for water environment appropriately. This paper takes the data from 2000 to 2010 to introduce the model and show both establishment and application, detailed procedures of which are as follows.

Model establishment

Screening sub-model

Initially, from the point of view of system, application and operability, several risk indexes are chosen to examine the water quality condition in Taihu basin, including: C_1 – chemical oxygen demand (COD_{Cr}); C_2 – 5-day biochemical oxygen demand (BOD_5); C_3 – ammonia nitrogen (NH_3-N); C_4 – total phosphorus (TP); C_5 – total nitrogen (TN); C_6 – dissolved oxygen (DO); C_7 –- chlorophyll a (chl-*a*); C_8 – permanganate index; C_9 – water temperature; C_{10} – pH value. This selection is based on Chinese environmental quality standards for surface water (GB3838–2002) and lake trophic status evaluation standards for water resources planning, as well as the representativeness and correlation of influencing index for water quality, taking the actual area data into consideration. Secondly, it invites experts to make a comparison in evaluation importance between every two indexes mentioned above to establish the fuzzy complementary judgement matrix. Thirdly, it uses the fuzzy analytic hierarchy process based on the accelerating genetic algorithm to obtain the indexes weight of matrix. After calculation, the consistency index coefficient value of each matrix is less than 0.2. So they have satisfactory consistency, and the calculated index weight is acceptable, as shown in Table 1 which illustrates that the ratios of standard deviation to

Table 1 | The calculated results of water quality assessment index system on AGA-FAHP in the plain area of Taihu basin

Weight eigenvalue	COD_{Cr}	BOD_5	NH_3-N	TP	TN	DO	ch1-*a*	Permanganate index	Water temperature	pH value
Mean value	0.113	0.121	0.126	0.117	0.120	0.103	0.075	0.094	0.072	0.059
Standard deviation	0.021	0.033	0.045	0.036	0.027	0.025	0.019	0.023	0.016	0.022
Standard deviation/Mean value	0.186	0.273	0.357	0.308	0.225	0.243	0.253	0.245	0.222	0.373

mean value for each index weight are small in most cases. This suggests that the experts' opinions are relatively concentrated. Therefore, six indexes with a relatively larger mean value are chosen to constitute the risk evaluation index system for water quality in Taihu basin, such as COD_{Cr}, BOD_5, NH_3-N, TP, TN, DO, etc.

Prediction and evaluation sub-models

Firstly, in this paper clustering samples are divided into five types in the plain area of the Taihu basin, namely $N = 5$. Secondly, assuming that the precipitation level for the flood season in 2000 is small and the condition is unfavorable to water allocation, this paper sets all the index data on January as the initial concentration with reference to actual data in 1999, and takes the Taihu basin model to simulate the water quality index of all sub-areas in 2000. For data from other years, from 2001 to 2009, the process can be repeated by analogy. Thirdly, the actual and simulated data of each sub-area from 2000 to 2009 are used to establish the dynamic risk assessment model for water quality on the projection pursuit cluster principle, where $n = 15$, $m = 6$. The concrete process is described as follows. On the basis of water quality evaluation index system, it puts the time-space stereo data into the evaluation model to be treated in dimensionless consistency, and turns all indexes to be of the type that the less, the better. Then it applies the accelerating genetic algorithm to optimize objective function in order to get a vector in the optimal projection direction: $d_d(a) = (0.432, 0.375, 0.413, 0.411, 0.305, 0.427)$. Finally, it brings this vector above into the index data of each year to get a risk assessment in the Taihu basin. Clustering results are standardized in interval from 0 to 5. Risk types are described as I, II, III, IV and V type, respectively. The first sub-area is taken as an example, the results of which are given in Table 2.

Then, the process mentioned above is repeated with the actual and simulated data in other sub-areas from 2000 to 2009, the results of which are shown in Table 3. Table 3 shows that there is a little difference in projection eigenvalues between the simulated and actual data, the relative errors are all less than 8%, cluster results which are calculated with simulated data conform to the actual, and the correction rate is very high. Hence, a dynamic risk assessment model based on the projection pursuit cluster principle can reasonably assess risk with index data to some extent which can be used to study risk assessment in water quality.

Model application

On the basis of the water quality assessment index system in the plain area of the Taihu basin using the screening

Table 2 | The comprehensive risk assessment from 2000 to 2009 in the first sub-area of the Taihu basin

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Projection eigenvalue	Clustering results
Actual	2.127	2.277	2.449	2.887	3.262	3.410	3.635	4.134	4.012	3.893	2.794	III
Simulated	2.132	2.298	2.513	2.890	3.301	3.413	3.710	4.142	4.103	3.905	2.811	III

Table 3 | The comprehensive risk assessment from 2000 to 2009 in the plain area of the Taihu basin

Sub-area		1	2	3	4	5	6	7	8
Projection eigenvalue	Actual	2.794	3.802	4.021	4.234	4.213	4.234	4.012	3.673
	Simulated	2.811	3.953	4.124	4.221	4.356	4.107	4.141	3.549
Clustering results		IV	V	V	V	V	V	IV	III
Sub-area		**9**	**10**	**11**	**12**	**13**	**14**	**16**	
Projection eigenvalue	Actual	4.134	3.879	4.011	4.234	4.102	4.373	2.234	
	Simulated	4.287	3.791	4.139	4.391	4.001	4.434	2.363	
Clustering results		IV	V	V	V	V	III	V	

Table 4 | The risk assessment in the plain area of the Taihu basin in 2010

Sub-area	1	2	3	4	5	6	7	8
Simulated	3.893	3.717	4.141	4.213	4.427	4.218	4.171	3.508
Sub-area	**9**	**10**	**11**	**12**	**13**	**14**	**16**	
Simulated	4.064	4.128	4.076	4.452	4.207	4.396	2.749	

Table 5 | The comprehensive risk assessment in the plain area of the Taihu basin from 2000 to 2010

Sub-area	1	2	3	4	5	6	7	8
Projection eigenvalue	2.890	3.875	4.221	4.219	4.362	4.116	4.145	3.538
Clustering results	III	IV	V	V	V	V	V	IV
Sub-area	**9**	**10**	**11**	**12**	**13**	**14**	**16**	
Projection eigenvalue	4.279	3.926	4.131	4.401	4.103	4.425	2.371	
Clustering results	V	IV	V	V	V	V	III	

sub-model, this paper uses the prediction and evaluation sub-models to calculate risk assessment value in 2010. Table 4 shows that risk trend of many sub-areas in 2010 is increasing, but only a few decline compared with the comprehensive risk from 2000 to 2009, and the fluctuation range is small on the whole. The results mentioned above combine with the simulated condition from 2000 to 2009 in order to get the comprehensive risk assessment in the plain area of the Taihu basin from 2000 to 2010. Table 5 shows that there is an obvious change in the comprehensive risk from 2000 to 2010 compared with the one from 2000 to 2009, whose trend relates to the condition in 2010. It can be considered that the first and 16th will be in III type, the second, eighth and 10th are in IV type, and others are in V type from 2000 to 2010. Thus, the proposed model may not only assess water quality risk in one year on a microscopic scale, but also give a description on risk condition in one time on the macroscopic scale.

Regulation countermeasures

The comprehensive assessment results from 2000 to 2010 demonstrate that the water pollution trend in the plain area of the Taihu basin is increasing to a higher warning state. Especially in 2007, the risk state of water quality starts turning to a higher type. This condition matches well with the outbreak of blue algae in May 2007. According to the situation calculated by the screening, prediction and evaluation model in 2010, the system will develop to become very polluted. If this situation were not controlled efficiently, it is possible that the local sustainable utilization of water resources and development of economic society would be harmed significantly. In order to change the trend of water pollution, project measures and non-project measures should be adopted in six facets to properly adjust water resources strategy for development and utilization. Firstly, the responsibility of local government on water pollution protection needs to be made clear, with the production of regulations to strictly reduce the total quantity of pollutant emission on the control index standard, and make no approval for any new project that may increase the quantity of major pollutants which may result in the exceeding of acceptable levels. Secondly, supervision law should be strengthened with regard to pollution sources from industry and the wastewater treatment plant of the city, carry out the discharge permit comprehensively, forbid sewage without permit or not on the discharge permit rule, and determine the principle of the over standard and illegal investigation at the same time. Thirdly, the protection range and protection measures regarding drinking water sources should be increased. Fourthly, increase the regulations in the grades of water pollution accidents, the emergency command institute, the emergency plan compilation, the notification and disposal of accident report, results monitoring, and information publishing and so on. Fifthly, it must complementarily comply the construction planning of treatment facilities in water pollution, and close enterprises who are heavy polluters or have no pollution prevention measures. Sixthly, increase the fine amount on common illegal behavior, and serve appropriate punishment on severe polluters, such as governance within a set time limit, restriction of production and excretion, stopping production and rectifying, or closure, punishment of public security management and investigating criminal responsibility, in order to solve the problem that the financial cost of law-abiding is high and law-breaking is low.

CONCLUSIONS

The dynamic risk assessment model for water quality based on the projection pursuit cluster principle combines projection pursuit cluster thinking with a comprehensive dynamic assessment method, which adequately brings the prominent advantage into play where projection pursuit treats with multidimensional data, completely makes use of data self-characteristics to cluster samples, avoids man-made influence on results of risk-area division due to experts' knowledge, and fully grasps spatial and temporal distribution of water quality risk, that is a physically based model with reasonable structure. Regarding establishment, firstly, the model uses the screening sub-model to obtain a risk evaluation index system for water quality. Secondly, it estimates the development tendency of the water quality risk index by selecting different prediction sub-models with actual conditions from the study area. Thirdly, it uses the projection pursuit cluster principle and comprehensive dynamic evaluation method to establish the dynamic assessment sub-model in order to study water quality assessment. These three sub-models complement each other and are indispensable. Regarding application, the model objectively takes a particular time as the appraisal time spot, qualitatively and quantitatively evaluates the possibility of water quality risk and the corresponding level in some future period, comprehensively analyzes the area's water quality risk development tendency, and strictly compartments water quality risk zoning. Moreover, with the increase and extension of historical data, the model continuously updates information in spatial and temporal scale, dynamically unfolds the entire process of water quality risk assessment. In this paper, the model is applied to the plain area of the Taihu basin. With model training and performance evaluation, it is shown that the developed model is simple in implementation, has the satisfactory stability, strong objectivity, and its results fit the actual situation well. So, the model can be continuously used to study assessment work of water quality. And therefore it may assist in planning schemes and emergency planning of water resources regulation, which not only breaks new ground for risk-area division analysis and development trend of risk assessment, but also provides a new idea for water pollution research.

ACKNOWLEDGEMENTS

The authors would like to thank the support of the Major Special Technological Program of Water Pollution Control and Management (Program No. 2009ZX07106-001) and the National Natural Science Funds of China (No. 51079037). The authors also want to thank the anonymous referees for their helpful suggestions and corrections on the earlier draft of our paper according to which we improved the content.

REFERENCES

Chen, X. H., Jiang, T. & Chen, J. H. 2007 *Water Environment Evaluation and Planning*. China WaterPower Press, Beijing, pp. 1–15.

Cheng, W. H., Wang, C. H. & Zhu, Y. 2006 *Taihu Basin Model*. Hohai University Press, Nanjing, pp. 11–36.

Gao, Z. J., Tian, H. & Zhang, C. R. 2007 A summary of water environment assessment. *Journal of Shandong University of Science and Technology (Natural Science)* **26**, 16–18.

He, R. M. 2007 *Study on the Theory and Evaluation Method of Regional Water Environment Carrying Capacity*. Hohai University Press, Nanjing, pp. 77–98.

Jin, J. L., Hong, T. Q. & Wang, W. S. 2007a Entropy and FAHP based fuzzy comprehensive evaluation model of water resources sustaining utilization. *Journal of Hydroelectric Engineering* **26**, 22–28.

Jin, J. L., Wu, K. Y. & Li, J. Q. 2007b Entropy coupling method of evaluating Chaohu Lake water quality security using

correspondence factor analysis and projection pursuit. *Journal of Sichuan University (Engineering Science Edition)* **39**, 7–13.
Jin, J. L., Wu, K. Y. & Wei, Y. M. 2008 Connection number based assessment model for watershed water security. *Journal of Hydraulic Engineering* **39**, 401–409.
Li, L. W. & Shi, Z. 2007 Experimental study on purification of slightly-polluted surface water by constructed rapid infiltration system. *Journal of Hydraulic Engineering* **38**, 417–426.
Li, R. Z., Hong, T. Q. & Jin, J. L. 2007 Research on fuzzy risk assessment model for river water quality. *Journal of Wuhan University of Technology* **29**, 43–46.
Wang, S. J. & Li, Y. Q. 2007 A dynamic cluster model based on projection pursuit with its application to climate zoning. *Journal of Applied Meteorological Science* **18**, 722–726.
Wang, S., Yang, S. L. & Hu, X. X. 2008 Combinational evaluation method based on projection pursuit algorithmn. *China Engineering Science* **60**, 60–64.
Wang, H., Qiu, Y. Q. & Jia, Y. W. 2010a Development course and tendency of water resources assessment. *Journal of Beijing Normal University (Natural Science Edition)* **46**, 274–277.
Wang, J., Cai, S. H., Yang, Y. W. & Li, R. J. 2010b Study on comprehensive grade evaluation of river health to Yihe River based on projection pursuit of entropy. *Journal of Water Resources and Water Engineering* **21**, 120–123.
Wu, K. Y., Jin, J. L. & Wei, Y. M. 2009 Intelligent integrated model for forewarning evaluation of watershed water security. *Advance in Water Science* **20**, 518–525.
Yang, X. H., Yang, Z. F. & Li, J. Q. 2004 A genetic projection pursuit interpolation model for comprehensive assessment of water quality. *Engineering Science* **22**, 66–71.
Yao, Y. & Ni, Q. 2009 A projection pursuit classification model based on K-Means dynamic cluster. *Journal of Nanjing Normal University (Natural Science Edition)* **32**, 16–20.
Yao, Z. M. & Zhang, J. Y. 2009 Research progress of water resources assessment. *Water Resources Research* **30**, 16–18.
Zhang, C. & Dong, S. H. 2009 A new water quality assessment model based on projection pursuit technique. *Journal of Environmental Sciences* **21**, 154–157.
Zhao, J. 2007 Research on water resources value assessment in plain area of taihu basin. Thesis. Hohai University, Nanjing.
Zhao, J. 2011 Study on the risk assessment of regional flood disaster and forewarning system. Thesis. Hohai University, Nanjing.

First received 10 October 2010; accepted in revised form 20 April 2011. Available online 16 March 2012